

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 14:07:17 ; Search time 15810.8 Seconds
(without alignments)
2705.363 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggctgaggtccactgtgct.....aaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

1	1973	96.5	2170	9	AF176838	AF176838 Homo sapi
2	1774.6	86.8	3278	9	AF246718	AF246718 Homo sapi
c	3	1753.6	85.8	157337	9	AC009163 Homo sapi
	4	1753.6	85.8	194832	9	AC025287 Homo sapi
	5	1742.6	85.3	159072	2	AC026419 Homo sapi
	6	1719	84.1	71503	9	AF219991 Homo sapi
	7	1248.6	61.1	1462	9	AF176839 Homo sapi
	8	1101	53.9	3786	9	AF280086 Homo sapi
	9	1052	51.5	2544	9	AF219990 Homo sapi
c	10	1046.4	51.2	159072	2	AC026419 Homo sapi
	11	1046.4	51.2	208185	2	AC009105 Homo sapi
	12	947	46.3	1647	6	AX327330 Sequence
	13	751.2	36.8	1989	10	AF176840 Mus muscu
	14	750.4	36.7	1740	10	AF176841 Mus muscu
	15	719.6	35.2	139252	2	AC095664 Rattus no
c	16	571.2	27.9	90256	9	AP001582 Homo sapi
	17	568	27.8	215647	2	AC068591 Homo sapi
c	18	556.8	27.2	174187	2	AP002792 Homo sapi
c	19	542.6	26.5	137499	2	AC015931 Homo sapi
	20	400	19.6	2032	9	AF131235 Homo sapi
	21	398.4	19.5	1333	9	AF149783 Homo sapi
	22	398.4	19.5	1992	9	AF280088 Homo sapi
	23	398.4	19.5	2011	9	AK026635 Homo sapi
	24	398.4	19.5	183228	9	AC010547 Homo sapi
	25	358.8	17.6	2201	10	AF109155 Mus muscu
	26	354	17.3	1926	10	AF131236 Mus muscu
	27	235.2	11.5	2156	6	AR071396 Sequence
	28	235.2	11.5	2156	6	E14937 Human mRNA
	29	235.2	11.5	2731	9	AB012192 Homo sapi
	30	235.2	11.5	6961	9	AB017915 Homo sapi
	31	235.2	11.5	178251	9	AC073370 Homo sapi
	32	235.2	11.5	196465	9	AC022392 Homo sapi
c	33	228.8	11.2	98392	2	AL590620 Homo sapi
c	34	228.8	11.2	172837	9	AL450304 Human DNA
c	35	226.2	11.1	103567	9	HS124C6 Homo sapi
c	36	224.4	11.0	54666	9	AC073487 Homo sapi
	37	223	10.9	176895	2	AC097065 Homo sapi
	38	222	10.9	140210	2	AC002993 Homo sapi
c	39	221.6	10.8	183974	9	AC093421 Homo sapi
c	40	220.8	10.8	151068	9	AL133405 Human DNA
c	41	220.6	10.8	148463	2	AC012140 Homo sapi
	42	220.6	10.8	207408	2	AC092872 Pan trogl
	43	220.6	10.8	209476	2	AC021027 Homo sapi
c	44	220.4	10.8	57573	2	AC087509 Homo sapi
	45	220.4	10.8	180317	2	AC025590 Homo sapi

ALIGNMENTS

RESULT	1
AF176838	AF176838
LOCUS	2170 bp mRNA linear PRI 22-SEP-1999
DEFINITION	Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase (I-GlcNAc-6-ST) mRNA, complete cds.
ACCESSION	AF176838
VERSION	AF176838.1
KEYWORDS	GI:5917705
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 2170)
TITLE	Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S. Cloning and characterization of a mammalian N-acetylglucosamine-6-sulfotransferase that is highly restricted to intestinal tissue
JOURNAL	Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)
MEDLINE	99423499
REFERENCE	2 (bases 1 to 2170)
AUTHORS	Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
TITLE	Direct Submission
JOURNAL	Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,

QY 1037 ttogaggacctggcgaggagccgctggcagagatcgcgcactctacgccttcacggc 1096
|||||
Db 2279 TTCGAGGACCTGGCGGGAGCCGCTGGCAGAGATCGCGCACTCTACGCCTTCACGGC 2338

QY 1097 ctgacctcacgccacagctcgaggcctggatccacaacatcacccacgggtcggggatc 1156
|||||
Db 2339 CTGACCCCTCAGCCACAGCTCGAGGCCTGGATCCACAACATCACCCACGGGTGCGGGATC 2398

QY 1157 ggcaagccaatcgaggccttcacatacttcgtctaggaatgcgcgaacgtctccaggcc 1216
Db 2399 GGCAAGCCAATCGAGGCCTTCACATACTTCGTCTAGGAATGCGGCAACGTCTCCAGGCC 2458

QY 1217 tggcgccacgcgttgcccttcactaagatccctgcgcgtgcaggaggtgtgcgcggcgcg 1276
Db 2459 TGGCGCCACGCGTGTGCCCTTCACATAAGATCCTGCGGTGCAGGAGGTGTGCGCGCGCGG 2518

QY 1277 ctgcagctgctgggtaccggccctgtgtactctcgggaccagcagcgtgacctcaccctg 1336
Db 2519 CTGCAGCTGCTGGGCTACCGGCCCTGTGTACTCTGCGGACCAGCAGCGTGACCTCACCCCTG 2578

QY 1337 gatctggtgctgccagagagcccgaaactgtgtgcgcagccacttcagctgggcacgtgcctgactgagaactc 1396
Db 2579 GATCTGGTGTGCTGCCACGAGGCCAGACCACCTTCAGTTGGCATCGGCTGCCTGACTGAGAACTC 2638

QY 1397 tgggccttagagcaagcccccgaactgtgtgcgcagggccaggaagcgcactgcatgttg 1456
Db 2639 TGGGCCTTAGAGCAGGCCCCGAACTGTGGTCGCGAGGCCAGGAGCGGACTCGATGGTGG 2698

QY 1457 aaaaggagctggggcgcatggggaa-caggtccctactatcaacccgggagtttgggtcc 1515
Db 2699 AGAGGAGCTGGGGCGCATGGGGAAGCAGGTCCCTACTATCAACCGGGAGTTGGGGTCC 2758

QY 1516 tccctgaagtaagcaaggactgcacgtttcttctctcctgattctcgttttcccttg 1575
Db 2759 TCCCTGAAGTAGGCAAGGACTGCACGTTTCTTCTCTCTCTGATTCTCGGTTTCTCTTG 2818

QY 1576 agtctctggagctgccttctcatcaggtgcactcttcatggaaaagcaactcttgcccc 1635
Db 2819 AGTCTTCTGGAGTGCCTTCTCATCAGGTGCCTCTTCATGG-AAAGCAACTCTTGCCCC 2877

QY 1636 tacctcttctggcgaggaagtaagtactgctaaattaaattaaatgtgtgccaggcc 1695
Db 2878 TACCTCTTCTGGCGCAGGGAGTAAGTTACTGCTAAATTAATTAATGTGTGCCAGGCC 2937

QY 1696 ggtggtggtgcctgctgtaatccagcattttgagaggtgagcggtgagcggtgatcac 1755
Db 2938 GGTGCGGTGGCTCATGCCCTGTAATCCAGCATTTTGAGAGGCTGAGGCGGTGGATCAC 2997

QY 1756 ctgaggtcagga-ttcaaaaccagcctggccaacatagtgaaacccctctctactaaaa 1814
Db 2998 CTGAGGTCAAGAGTTCGAAACCAGCCTGGCCAACATAGTGAACCCCTCTCTACTAAA 3057

QY 1815 atgcataaattagtcgcgcgtggtggcacactcctgtaatccccagctacttaggagctg 1874
Db 3058 ATGCATAAATTAGTCCGGCGTGGTGGCACACTCCTGTAAATCCCACTACTTAGGAGGCTG 3117

QY 1875 aggtgggaaaaatcacttggactccaaagggtggaggttgcagtaagctgaaatcatgccac 1934
Db 3118 AGGTGGGAGAATCATTGGACTCCAGAGGTGGAGGTTGCAGTAAGCTGAGATCATATGCCAC 3177

QY 1935 tgcacctagcttgggtggcaaaagcaaaactctatcaaaaaataataataatttgtt 1994
Db 3178 TGCACCCCTAGCTTGGGTGGCAGAGCAAGACTCTATCAAAAAAATAAGTAATAAATTTGTT 3237

QY 1995 caaaagtcctgccgcaaaaaaataaaaaaataaaaaa 2035
Db 3238 CAAAAGTCCGAAAAAATAAAAAAATAAAAAAATAAAAAA 3278

RESULT 3
AC009163/c
LOCUS AC009163 157337 bp DNA linear PRI 02-NOV-2001

DEFINITION Homo sapiens chromosome 16 clone RP11-77K12, complete sequence.
ACCESSION AC009163
VERSION AC009163.5 GI:16596526
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 157337)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 157337)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
AUTHORS 3 (bases 1 to 157337)
TITLE Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 157337)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 2, 2001 this sequence version replaced gi:7689976.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

FEATURES
Location/Qualifiers
Source
1..157337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-77K12"
BASE COUNT 45295 a 36632 c 36308 g 39102 t
ORIGIN

Query Match 85.8%; Score 1753.6; DB 9; Length 157337;
Best Local Similarity 99.1%; Pred. No. 3.2e-268;
Matches 1795; Conservative 0; Mismatches 14; Indels 3; Gaps 3;

QY 197 cggtagccccccagccgccccgcacgtggtgctgcacggttctccagcaagacagtgacagtg 256
|||
Db 144606 CGAAGGCCCCCAGCCGCCCATGTGGCTGCACGGTTCTCCAGCAAGACAGTGACAGTG 144547

QY 257 ctctctctggcacagaccctgcctcctgctcttcttcacatctcccggccaggccctca 316
|||
Db 144546 CTCCTCCTGGCACAGACCACCTGCCTCTCTCTCTCATCATCTCCCGCCAGGCCCTCA 144487

QY 317 tccccagccggcgaggagatcgtgtgacgtgctggtgctgctcctcgtggtcggtcggtcg 376
|||
Db 144486 TCCCCAGCCGGCGGCGGAGGATCGTGTGACGTGCTGCTGTCTCTCTGTCGTCGCGC 144427

QY 377 tcataccttcttgggcccagctcttcagccagcaccaccccgacgtcttctacatgatggagccc 436
|||
Db 144426 TCATCCTTCTTGGGCCAGCTCTTCAGCCAGCACCCCGACGCTCTTCTACCTGATGGAGCCC 144367

QY 437 gcgtggcatgtgtggaccaccctgtcgcaggggcagcgcggaacgctgcacatggccgtg 496
|||||
Db 144366 GCGTGGCATGTGTGGACACACCTGTTCGCAGGGCAGCGCGGAACGCTGCACATGGCCGTG 144307
QY 497 cgcgaacctgatgcgctctatctttttgtgcgacatggaacgtgtttgatgcctacatgcc 556
|||||
Db 144306 CCGCACCTGATCGGCTCTATCTTTTGTTCGACATGGACGTGTTGATGCCTACATGCCA 144247
QY 557 cagagccgaacctgtccgccttttcccgagggcaccatcagcaagcaggccgcgctgtgctgcgcg 616
|||||
Db 144246 CAGAGCCGAAACCTGTCCGCCTTTTCAACTGGGCAACGAGCGCGCGCTGTGCTCGCGG 144187
QY 617 cccgcctgcagcgccctttcccgagggcaccatcagcaagcaggacgtatgcaagacactg 676
|||||
Db 144186 CCGGCTGCAGCGCCTTTCGCCGAGGCACCATCAGCAAGCAGGACGTATGCAAGACACTG 144127
QY 677 tgcacgcgcagccattcagccctggcccgggaggccctgcgcctcctacagccacagtgtg 736
|||||
Db 144126 TGCACGCGGCAGCCATTTCAGCCTGGCCCGGGAGGCGCTGCCGCTCCTACAGCCACGTGCTG 144067
QY 737 ctcaaggaggtgcgctcttctcaacctgcaggtgctctacccgctgctcagcgacccccgcg 796
|||||
Db 144066 CTCAAGGAGGTGCGCTTCTTCAACCTGCAGGTGCTCTACCGGCTGCTCAGCGACCCCGCG 144007
QY 797 ctcaacctgcgcctcgtgcacctggtgcgcgaccccgcgccgctgctgcgctcccgggag 856
|||||
Db 144006 CTCAACCTGCGCATCGTGCACCTGGTGCAGACCCCGCGGCGGTGCTGCGCTCCCGGGAG 143947
QY 857 gcggcgggcccgatactggaacgcgcgacaaacggcatcgtctgggcaccacggcaagtgg 916
|||||
Db 143946 GCGGCGGGCCCGATACTGCGCACGCGACAAACGGCATCGTGTGGGCACCAACGGCAAGTGG 143887
QY 917 gtggaggccgacccctcaacctgcgcctgattcgcgaggtgtgcgcgagccacgtgcgcac 976
|||||
Db 143886 GTGGAGGCGCGACCCCTCACCTGCGCCTGATTTCGCGAGGTGTGCGCGCAGCCACGTGCGCATC 143827
QY 977 gcgcaggccgcacacctcaagccgcaccccttccctgcgcgcgcgtacccgctggtgcgc 1036
|||||
Db 143826 GCGAGGCGCGCACACTCAAGCGGCCACCCCTTCCTGCGCGGCGGTACCGCCTGTTGCCG 143767
QY 1037 ttgcaggacctggcgcgggagcgcgctggcagagatccgcgcactctacgccttcacccgc 1096
|||||
Db 143766 TTCCAGGACCTGGCGCGGGAGCGCGTGGCAGAGATCCCGCGACTCTACGCCCTTCACCCGC 143707
QY 1097 ctgacctcagccacagctcaggccctggatccacaacatcaccacaggggtcggggatc 1156
|||||
Db 143706 CTGACCCCTCAGCCACAGCTCGAGGCGCTGGATCCACAACATCACCCACGGGTTCGGGGATC 143647
QY 1157 ggcaagccaatcgaggccttccatacttcgtctaggaatgcgcgaacgtctccaggcc 1216
|||||
Db 143646 GGCAAGCCAAATCGAGGGCCTTCCATACTTCTGCTAGGAATGCGCGCAACGTCTCCCAGGCC 143587
QY 1217 tggcgccacgcgttgcccttcaataagatccctgcgcgtgcaggaggtgtgcgcgcgcgcg 1276
|||||
Db 143586 TGGCGCCACCGCTTGCCCTTCACTAAGATCCTGCGCGGTGCAGGAGGTGTGCGCGCGCGCG 143527
QY 1277 ctgcagctgctgggctaccgcccctgtgtactctcgcgaccagcagcgtgacctcacccctg 1336
|||||
Db 143526 CTGCAGCTGCTGGGCTACCGGCGCTGTGTACTCTGCGGACCAAGCAGCGGTGACCTCACCCCTG 143467
QY 1337 gatctggtgctgccacgagcccgagaccacttcagctgggcctgcctgactgagaactc 1396
|||||
Db 143466 GATCTGCTGCTGCCACGAGGCCCGACCACTTCAGCTGGGCATCGCCTGACTGAGAATC 143407
QY 1397 tggcccttagagcaagccccgaactgtggtgcgccaggcccggaagcgcactgcagtgtg 1456
|||||
Db 143406 TGGGCGCTTAGAGCAGGCCCGCACTGTGGTGCGCCAGGCCCGCAGGCGGACTGCATGGTGG 143347
QY 1457 aaaaggagctggggcgcatggggaa-caggtccctactatcaacggggagtttggggctcc 1515
|||
Db 143346 AGAGGGAGCTGGGGCGCATGGGGAAGCAGGTCCTACTATCAACCGGGAGTTTGGGGTCC 143287
QY 1516 tccccgaagtaagcaaggactgcaggtttcttctcctcctgattctcggttttcctttg 1575

Db 143286 TCCCTGAAGTAGGCAAGGACTGCACGTTTCTTTCTCTCTCGATTCTCGGTTTTCCTTTG 143227
QY 1576 agtcttctggagctgccttctcactcaggtgcactcttctatggaagcaactcttgcgcc 1635
Db 143226 AGTCTTCTGGAGTGCTTCTCATCAGGTGCACTCTTTCATGG-AAAGCAACTCTTGCCCC 143168
QY 1636 tacctcttctggcgccagggagtaagtactgctaaaaattaaatgtgtgccagggcc 1695
|||||
Db 143167 TACCTCTTCTGGCGCAGGGAGTAAGTTACTGCTAAATTAAATTAAATGTGTGCCAGGCC 143108
QY 1696 ggtgcggtggctcactgcctgtaatccccagcatttttgagaggtgcaggcggtggatcac 1755
Db 143107 GGTGCGGTGGCTCATGCGCTGTAATCCCAGCATTTTGAGAGGCTGAGCGGGTGGATCAC 143048
QY 1756 ctgaggtcagga-ttcaaaaaaccagcctggccaacatagtgaaacccctctctactaaaa 1814
Db 143047 CTGAGGTCAGGAGTTCGAAACCCAGCTGGCCAACATAGTGAAACCCCTCTCTACTAAAA 142988
QY 1815 atgcaaaaaattagtcggcgctggtggcacactcctgtaatcccagctacttaggagctg 1874
Db 142987 ATGCAAAAAATTAGTCCGGCGTGGTGGCACACTCCTGTATATCCCACTACTTAGGAGGCTG 142928
QY 1875 agtgggaaaaatcaacttggaactccaaagtggaaggtgcagttgcagtaagtgaaatcatgccac 1934
Db 142927 AGGTGGAGAAATCACTTGGACTCCAGAGGTGGAGGTTCAGTAAGTGAGATCATGCCAC 142868
QY 1935 tgcacctagcttgggtggcaagcaaaactctatcaaaaaataaataaaatttgtt 1994
Db 142867 TGCACCTAGCTTGGGTGGCAGAGCAAGACTCTATCAAAAAAATAAGTAATAAATTTGTT 142808
QY 1995 caaaagtccctgc 2006
Db 142807 CAAAAGTCTGC 142796

RESULT 4

AC025287
LOCUS AC025287 194832 bp DNA linear PRI 26-JAN-2002
DEFINITION Homo sapiens chromosome 16 clone RP11-490B18, complete sequence.
ACCESSION AC025287
VERSION AC025287.8 GI:18376863
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jan 26, 2002 this sequence version replaced gi:17976465.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu
Quality: Phrap Quality >=40 99.4% of Sequence;
Estimated Total Number of Errors is 0.6.

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159072)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 159072)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 20, 2001 this sequence version replaced gi:7711893.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 671130, H460
Center clone name: CITB-HL_2113H21

Summary Statistics
Consensus quality: 140723 bases at least Q40
Consensus quality: 148960 bases at least Q30
Consensus quality: 151290 bases at least Q20
Estimated insert size: 98000; pulse field gel estimation
Estimated coverage: 7.39 in Q20 bases; pulse field gel estimation
Quality coverage: 4.59 in Q20 bases; sum-of-contigs estimation.
Quality coverage: 4.59 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1084: contig of 1084 bp in length
* 1085 1184: gap of unknown length
* 1185 2395: contig of 1211 bp in length
* 2396 2495: gap of unknown length
* 2496 4280: contig of 1785 bp in length
* 4281 4380: gap of unknown length
* 4381 6158: contig of 1778 bp in length
* 6159 6258: gap of unknown length
* 6259 8612: contig of 2354 bp in length
* 8613 8712: gap of unknown length
* 8713 11625: contig of 2913 bp in length
* 11626 11726: gap of unknown length
* 11726 13257: contig of 1531 bp in length
* 13257 13357: gap of unknown length
* 13357 16479: contig of 3123 bp in length
* 16480 16579: gap of unknown length
* 16580 19343: contig of 2764 bp in length
* 19344 19443: gap of unknown length
* 19444 23596: contig of 4153 bp in length
* 23597 23697: gap of unknown length
* 23697 29181: contig of 5485 bp in length
* 29182 29281: gap of unknown length
* 29282 35079: contig of 5798 bp in length
* 35080 35179: gap of unknown length
* 35180 39189: contig of 4010 bp in length
* 39190 39289: gap of unknown length
* 39290 48543: contig of 9254 bp in length
* 48544 48643: gap of unknown length
* 48644 90448: contig of 41805 bp in length
* 90449 90548: gap of unknown length
* 90549 159072: contig of 68524 bp in length.
Location/Qualifiers
1. .159072
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2113H21"
/clone_lib="CalTech human BAC library D"
BASE COUNT 43207 a 36742 c 36517 g 41088 t 1518 others
ORIGIN

Query Match 85.3%; Score 1742.6; DB 2; Length 159072;
Best Local Similarity 99.0%; Pred. No. 1.7e-266;
Matches 1795; Conservative 0; Mismatches 14; Indels 4; Gaps 4;

QY 197 cgggtgccccagcgcgcgcgtgtggctgccacgggttctccagcaagacagtgacagtg 256
|||
Db 61379 CGAAGGCCCCAGCCGCCGCATGTGGTGCCACGGTCTCTCCAGCAAGACAGTACAGTG 61438

QY 257 ctctctctggcacagaccactgcctcctcctgtcttcttcatcatctcccggccagggccctca 316
|||
Db 61439 CTCCTCCTGGCACAGACCACCTGCCTCCTGCTCTTTCATCATCTCCCGCCAGGGCCCTCA 61498

QY 317 tccccagccggcggcgagatcgtgtgcacgtgtcgtgtgtcctcgttgcctggcgctgggc 376
|||
Db 61499 TCCCCAGCCGGCGGGGAGATCGTGTGCACGTGTGCTGTGCTCTCCTCGTGGCGCTCGGC 61558

QY 377 tcctcttcttgggcccagctcttccagccagcaccgccagcttcttacctgatggagccc 436
|||
Db 61559 TCATCCTTCTTGGGCCAGCTCTTCAGCCAGCACCCCGACGCTCTTACCTGATGGAGCCC 61618

QY 437 gcgtggcatgtgtggaccacccctgtcgcagggcgagcgccagcgcaacgctgcacatggccgtg 496
|||
Db 61619 GCGTGGCATGTGTGGACCACCTGTGTGCAGGGCGAGCGGCAACGCTGCACATGGCCGTG 61678

QY 497 cgcgacctgatgcgtctctatcttttgcgacatggagctgttgcctacatgccca 556
|||
Db 61679 CGGACCTGATGCGCTCTATCTTTTGTGCGACATGGACGTGTTTGTATGCCTACATGCCA 61738

QY 557 cagagccgaaaacctgtccgccttttcaactgggcaacgagcgcgctgtgctgcgcg 616
|||
Db 61739 CAGAGCCGAAACCTGTCCGCCCTTTTCAACTGGGCAACGAGCGCGCGTGTGCTCGCGG 61798

QY 617 ccgcctgcagcgcctttcccgaggcaccatcagcaagcaggacgtatgcaagacactg 676
Db 61799 CCGGCTGCAGCGCCTTTCCCGAGGGCACCATTACAGCAAGCAGGACGTTATGCAAGACACTG 61858

QY 677 tgcacgcggcagccattcagctgtggccgggagggcctgcgcctcctacagccacgtggtg 736
|||
Db 61859 TGCACGGCGGACGCCATTTCAGCTGGCCCCGGAGGCGCTGCCGCTCCTACAGCCACGTTGTTG 61918

QY 737 ctcaaggaggtgcgtcttcttcaacctgcaggtgtctctaccggctgtcagcgaccccgcg 796
Db 61919 CTCAAGGAGGTGCGCTTCTTCAACCTGCAGGTGCTTCTACCGCTGCTCAGCGACCCCGCG 61978

QY 797 ctcaacctgcgcacatcgtgcacgtggtgcgcgaccccgggcgctgtgcgtcccggaag 856
|||
Db 61979 CTCAACCTGCGCATCGTGCACCTGCTGCGCAGACCCCGGGCGGTGCTGCGTCCCGGAG 62038

QY 857 gggcgggcccgatactggcacgcgacacacggcatcgtgtggggcaccaacggcaagtgg 916
|||
Db 62039 GCGCGGGCCCCGATACCTGCGACCGACACGCGCATCGTGTGTTGGCACCAACGGCAAGTGG 62098

QY 917 gtggaggccgacctcactgcgcctgatttcgcgaggtgtgcgcgagccacgtgcgcac 976
|||
Db 62099 GTGGAGGCCGACCCCTCACCTGCGCCTGATTTCGGGAGGTGTGCCGAGCCACGTCGCCCATC 62158

QY 977 gccgaggccgccacactcaagccgcccaccttctcctgcgcgcccgtaccgcctgtgtgcgc 1036
|||
Db 62159 GCCGAGGCCGCCACACTCAAGCGGCCACCCCTTCTCTGCGCGCGCGTACCGCCTTGTGCGC 62218

QY 1037 ttcgaggacctggcgcgggagcgctggcagagatccgcgacctctacgccttcacccggc 1096
|||
Db 62219 TTCGAGGACCTGGCGGGGAGCGCGTGGCAGAGATCCGCGGCACTCTACGCTTTCACCGGC 62278

QY 1097 ctgacctcagccacagctcgagcgctggatccacacaatcacccacgggtcggggatc 1156
|||||
Db 62279 CTGACCCCTCAGCCACAGCTCGAGGCCCTGGATCCACAACATCACCCACGGGTGGGGATC 62338

QY 1157 ggcaagccaatcgagggccttccatacttcgttaggaatgcgcgaacgtctcccaggcc 1216
|||||
Db 62339 GGCAAGCCAAATCAGGCCCTTCCATACTTCGTCTAGGAATGCGCGCAACGTCTCCAGGCC 62398

QY 1217 tggcgccacgcttgcccttcactaagatcctgcgcgtgcaggaggtgtgcgcggcgcg 1276
|||||
Db 62399 TGGCGCCACGCGTGCCTTCACTAAGATCCTCGCGGTGCAGGAGGTGTGCGCGCGCGCG 62458

QY 1277 ctgcagctgclygggtaccggcctgtgtactctgcggaccagcagcgtgacctcaccctg 1336
|||||
Db 62459 CTGCAGCTGCTGGGTACCGGCTGTGTACTCTGCGGACCAGCAGCGTGACCTCACCCCTG 62518

QY 1337 gatctggtgtccacgagggcccagaccacttcagctgggcacgcctgcctgactgagaactc 1396
|||||
Db 62519 GATCTGGTGTGCCACGAGGCCAGCCACTTCAGCTCGGCAATCGGCTGACTGAGAACTC 62578

QY 1397 tgggccccttagagcaagccccgaactgtggtgcggcggccccaggaagcgaactgcactggtgg 1456
|||||
Db 62579 TGGGCCCTAGACGAGGCCCCCGAACTGTGGTGCGCCAGGCCAGGAGCGCACTGCATGTTGG 62638

QY 1457 aaaaggagctggggcgcatggggaa-caggtccctactatcaacccggagtttggggctcc 1515
|||
Db 62639 AGAGGGAGCTGGGGCGCATGGGGAAGCAGGTCCTACTATCAACCGGAGTTTGGGGTCC 62698

QY 1516 t-cccccttaagtaagcaaggactgcacgtttcttctctctcctgattctcgttttccctt 1574
|||||
Db 62699 TCCCCCTGAAGTAGGCAAGGACTGCACTGTCACGTTCTTCTCTCCTGATTCTCGGTTTCTCCTTT 62758

QY 1575 gagtcttgagctgccttctcatcaggtgcactcttcatggaaagcaactcttgcct 1634
|||||
Db 62759 GAGTCTTCTGGAGCTGCCTTCTCATCAGGTGCACCTTTTCATGG-AAAGCAACTCTTGCCC 62817

QY 1635 ctacctcttctggcgccagggagtaagttactgctaaattaaattaaatgtgtgccaggc 1694
|||||
Db 62818 CTACCTCTCTGGGGCGCAGGGAGTAAGTACTGTAAATTAAATTAATGTGTGCCAGGC 62877

QY 1695 cgggtgcggtggctcatgcctgtaatcccgagcattttgagaggtgcggcggtggatca 1754
|||||
Db 62878 CGGGTGGGTGGTCTCATGCCCTGTAATCCCAGCATTTTGAGAGGCTGAGGCGGGTGGATCA 62937

QY 1755 cctgaggtcagga-ttcaaaaaccagccctggccaaacatagtgaaacccctctctactaaa 1813
|||||
Db 62938 CCTGAGTTCAGGAGTTTCGAAACCAGCCTGGCCAAACATAGTGAACCCCTCTCTACTAAA 62997

QY 1814 aatgcaaaaaattagtcggcgctgggtggcacactcctgtaatcccagctacttaggagct 1873
|||||
Db 62998 AATGCAAAAATTAGTCCGGCGTGGTGGCACACTCTGTAAATCCAGCTACTTAGGAGGCT 63057

QY 1874 gagtggggaaaaatcacttggactccaaaaggttgagggttcagtaagctgaaatcatgcca 1933
|||||
Db 63058 GAGGTGGGAGAATCACTTGGACTCCAGAGGTGGAGGTTGCAGTAAGCTGAGATCATGCCA 63117

QY 1934 ctgcacactagcttgggtggcaagcaaaaactctatcaaaaaataattaataatttgt 1993
|||||
Db 63118 CTGCACCTAGCTTGGGTGGCAGAGCAAGACTCTATCAAAAAATAAGTAATAATTGT 63177

QY 1994 tcaaaagtccctgc 2006
|||||
Db 63178 TCAAAAGTCCCTGC 63190

RESULT 6
AF219991 AF219991 71503 bp DNA linear PRI 26-OCT-2000
LOCUS Homo sapiens intestinal N-acetylglucosamine-6-O-sulfotransferase
DEFINITION (CHST5) and corneal N-acetylglucosamine-6-O-sulfotransferase
(CHST6) genes, complete cds.
ACCESSION AF219991
VERSION AF219991.1 GI:11023147

KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 71503)
AUTHORS Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,T.,
Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,
Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.
TITLE Macular corneal dystrophy type I and type II are caused by distinct
mutations in a new sulphotransferase gene
JOURNAL Nat. Genet. 26 (2), 237-241 (2000)
MEDLINE 20472330
PUBMED 11017086
REFERENCE 2 (bases 1 to 71503)
AUTHORS Akama,T.O. and Fukuda,M.N.
TITLE Direct Submission
JOURNAL Submitted (29-DEC-1999) Glycobiology Program, The Burnham
Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
1. .71503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q22"
/clone="CITB-483K2"
358. .461
/rpt_family="L2"
/rpt_type=dispersed
complement(462. .771)
/rpt_family="Alu"
/rpt_type=dispersed
772. .945
/rpt_family="L2"
/rpt_type=dispersed
1169. .1338
/rpt_family="MIR"
/rpt_type=dispersed
complement(1553. .1804)
/rpt_family="Alu"
/rpt_type=dispersed
2616. .2916
/rpt_family="Alu"
/rpt_type=dispersed
3309. .3532
/note="similar to EST H12297 (EST cluster Hs.31147)"
3865. .4028
/note="similar to EST H12297 (EST cluster Hs.31147)"
4099. .4406
/rpt_family="Alu"
/rpt_type=dispersed
4793. .5250
/rpt_family="L2"
/rpt_type=dispersed
complement(5251. .5329)
/rpt_family="MIR"
/rpt_type=dispersed
complement(5768. .6080)
/rpt_family="Alu"
/rpt_type=dispersed
complement(6081. .6265)
/rpt_family="Alu"
/rpt_type=dispersed
6111. .6567
/note="similar to EST AA001593 (EST cluster Hs.6853)"
8035. .8350
/rpt_family="Alu"
/rpt_type=dispersed
10545. .11049
/note="similar to EST AI950923 (EST cluster Hs.6853)"
11043. .11195
/rpt_family="Alu"
/rpt_type=dispersed

```
repeat_region 11200. .11222
/rpt_family="Simple_repeat"
/rpt_type=tandem
repeat_region 11570. .11871
/rpt_family="Alu"
/rpt_type=dispersed
complement(12065. .12198)
/rpt_family="L1"
/rpt_type=dispersed
complement(12205. .12493)
/rpt_family="Alu"
/rpt_type=dispersed
complement(12813. .13110)
/rpt_family="Alu"
/rpt_type=dispersed
join(14185. .14228,16169. .16263,17712. .20821)
/gene="CHST5"
/product="intestinal N-acetylglucosamine
6-O-sulfotransferase"
14185. .20821
/gene="CHST5"
14681. .14713
/rpt_family="Simple_repeat"
/rpt_type=tandem
complement(14715. .14990)
/rpt_family="Alu"
/rpt_type=dispersed
complement(14995. .15288)
/rpt_family="Alu"
/rpt_type=dispersed
complement(15292. .15589)
/rpt_family="Alu"
/rpt_type=dispersed
15668. .15718
/rpt_family="Alu"
/rpt_type=dispersed
15977. .16098
/rpt_family="MIR"
/rpt_type=dispersed
complement(17298. .17433)
/rpt_family="MIR"
/rpt_type=dispersed
complement(18001. .18310)
/rpt_family="Alu"
/rpt_type=dispersed
complement(18420. .18591)
/rpt_family="MIR"
/rpt_type=dispersed
18732. .18767
/rpt_family="Simple_repeat"
/rpt_type=tandem
18967. .20202
/gene="CHST5"
/note="I-GlcNAc6ST; carbohydrate sulfotransferase 5;
expressed in the small intestine and colon"
/codon_start=1
/product="intestinal
N-acetylglucosamine-6-O-sulfotransferase"
/protein_id="AAG26326.1"
/db_xref="GI:11023148"
/translation="MGMRARVPKVAHSTRRPPAARMWLPFSSKTVTVLLAQTTCLL
LFITRPGPSSPAGGEDRVHVLVLSWSRSGSSFGLQLFQHPDVFIYLMPEAHVWTTL
SQGSAATHMAVRDLMRISIFLCMDVFDAYMPQSRNLSAFFNWATSRALCSPACSAF
PRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFNQLVLYPLLSDPALNLR
IVHLVRDPRAVLRSREAAGPILARDNGIVLGTNGKWEADPHLRILREVCRSVRIAE
AATLKPPFPFLGRYRLVRFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNIHSGSI
GKPIEAFTSSRNARNVSWAQRHALPFTKILRVQEVCAQALQLLGYRPVYSADQQRDL
TLDLVLPRGPDHFSWASPD"
20504. .20806
/rpt_family="Alu"
/rpt_type=dispersed
20869. .20961
/rpt_family="L1"
```

```
repeat_region 20962. .21200
/rpt_family="MER4-group"
/rpt_type=dispersed
21325. .21429
/rpt_family="MER4-group"
/rpt_type=dispersed
21430. .21722
/rpt_family="Alu"
/rpt_type=dispersed
21725. .22007
/rpt_family="Alu"
/rpt_type=dispersed
22008. .22040
/rpt_family="Simple_repeat"
/rpt_type=tandem
22041. .22313
/rpt_family="MER4-group"
/rpt_type=dispersed
22325. .22593
/rpt_family="Alu"
/rpt_type=dispersed
22594. .22949
/rpt_family="MER4-group"
/rpt_type=dispersed
23116. .23221
/rpt_family="Simple_repeat"
/rpt_type=tandem
23344. .23400
/rpt_family="MaLR"
/rpt_type=dispersed
23433. .23456
/rpt_family="Simple_repeat"
/rpt_type=tandem
complement(23457. .23720)
/rpt_family="Alu"
/rpt_type=dispersed
complement(23789. .24140)
/rpt_family="MER21-group"
/rpt_type=dispersed
24141. .24287
/rpt_family="MaLR"
/rpt_type=dispersed
complement(24308. .24420)
/rpt_family="MaLR"
/rpt_type=dispersed
complement(24467. .24528)
/rpt_family="MER1_type"
```

Query Match 84.1%; Score 1719; DB 9; Length 71503;
Best Local Similarity 98.5%; Pred. No. 9.4e-263;
Matches 1789; Conservative 0; Mismatches 20; Indels 8; Gaps 5;

```
QY 197 cgggtgccccagcgccgcgcatgtggctgccacggttctccagcaagacagtgacagtg 256
|| |||||
Db 19009 CGAAGGCCCCAGCGCGCCGCATGTGGCTGCCACGGTTCTCCAGCAAGACAGTGACAGTG 19068

QY 257 ctctctctggcacagaccacctgcctcctgcctcttccttcacatctctcccgccagggccctca 316
|| |||||
Db 19069 CTCCTCTGGCACAGACCACCTGCCTCCTGCTCTTCATCATCTCCCGGCCAGGGCCCTCA 19128

QY 317 tccccagcggcgcgaggatcgtgtgcacgtgctggtcgtcctcgtgctcctcgtggtcgtcgggc 376
|| |||||
Db 19129 TCCCCAGCGCGCGGAGGATCGTGTGCACGTGCTGGTGTCTCCTCGTGGCGCTCGGGC 19188

QY 377 tcctccttcttggcccagctcttcagccagcacccccagcagcttctacctgatggagccc 436
|| |||||
Db 19189 TCATCCTTCTTGGGCCAGCTCTTCAGCCAGCACCCCCGACCTCTTCTACCTGATGGAGCCC 19248

QY 437 gcgtggcatgtgtggaccacctgtgcaggggcagcgccagcgccacgctgcacatggccgtg 496
|| |||||
Db 19249 GCGTGGCATGTGTGGACCCCTGTTCGACGGGCGGCGGCAACGCTGCACATGGCCGTG 19308
```

QY 497 cgcgacctgatgcgctctatatctttttgtggacaatggacgtgttttgatgcctacaatgcca 556
|||||
Db 19309 CGCGACCTGATCGGCTCTATCTTTTGTGGACATGAGACGTGTTGATGCTACATGCCA 19368

QY 557 cagagccgaacctgtccgccttttccaactgggcaacagcgcgcgtgtgtcgcgcg 616
|||||
Db 19369 CAGAGCCGAACCTGTCCGCCCTTTTCAACTGGGCAACGAGCGCGCGCTGTCTCGCCG 19428

QY 617 cccgcctgcagccctttccccaggaccatcagcaagcaggacgtatgcaagacactg 676
|||||
Db 19429 CCGGCCTGCAGCGCCTTTCCCGAGGCACCATCAGCAAGCAGGACGTATGCAAGACACTG 19488

QY 677 tgcacggcgagccattcagccctggcccggaggccctgcgcctcctacagccacgtggtg 736
|||||
Db 19489 TGCACGGCGAGCCATTTCAGCCTGGCCCGGAGGCCCTGCCGCTCCTACAGCCACGTGGTG 19548

QY 737 ctcaaggagtgcgcttcttcaacctgcaggtgctctacctgcctgcctcagcagcccccg 796
|||||
Db 19549 CTCAAGGAGGTGCGCTTCTTCAACCTGCAGGTGCTCTACCCGCTGCTCAGCGACCCCGG 19608

QY 797 ctcaacctgcgcctcgtgcacctgtgtgcgcgacctgcgcgcctgcctgcctccccggag 856
|||||
Db 19609 CTCAACCTGCGCATCGTGCACCTGTGTGGCGGACCCGCGGCCGCTGCTCGGCTCCCGGGAG 19668

QY 857 gcggcgggcccgatactggcacgcgcgacaacggccatcgtgtcgtgggcaaccaacggcaagtgg 916
|||||
Db 19669 GCGGCGGGCCCGATACTGGCACGCGCACAAAGGCATCGTGTGGGCACCAACGGCAAGTGG 19728

QY 917 gtggaggccgacctcaacctgcacctgattcgcgaggtgtgcgcgagccacgtgcgcac 976
|||||
Db 19729 GTGGAGGCGGACCCCTCACTGCGCCCTTGATTCGCGAGGTGTGCCGCGAGCCACGTGCCGCATC 19788

QY 977 gccgaggccgcccacactcaagccgcacaccttccctgcgcgcgcgtacctaccgcctggtgcgc 1036
|||||
Db 19789 GCGGAGGCGGCCACACTCAAGCCGCGCACCCCTTCTCGCGGCGGCTACCGCCTGGTGCGC 19848

QY 1037 ttcgaggacctggcgcgggagccgctggcagagatccgcgcactctacgccttcaccgcgc 1096
|||||
Db 19849 TTCGAGGACCTGGCGCGGGAGCCGCTGGCAGAGATCCGCGCACTCTACGCTTCAACCGGC 19908

QY 1097 ctgacctcacgccacagctcgaggcctggatccacaacatcacccacgggtcggggatc 1156
|||||
Db 19909 CTGACCTCAGGCCACACAGCTCGAGGCTGGATCCACAACATCACCCACGGGTCGGGGATC 19968

QY 1157 ggcaagccaatcgaggccttccatacttgccttaggaatgcgcgcgaacgtctccccagcc 1216
|||||
Db 19969 GGCAAGCCAATCGAGGCCTTCCATACTTCGTCTAGGAATGCGCGCAACGCTCTCCAGGCC 20028

QY 1217 tggcgccaacgcttgcccttcaataagatccctgcgcgtgcagaggtgtgcgcgcgcgcg 1276
|||||
Db 20029 TGGCGCCACGCGTTGCCCTTCACTAAGATCCTGCGCGTGCAGAGGTGTGCGCGCGCG 20088

QY 1277 ctgcagctgctgggctaccggccctgtgtactctgcggaccagcagcgtgacctcacccctg 1336
|||||
Db 20089 CTGCAGCTGCTGGGCTACCGGCCTGTGTACTCTGCGGACCAAGCGTGACCTCACCCCTG 20148

QY 1337 gatctggtgctgccacgagcccagaccacttcagctgggcatcgccctgactgagaactc 1396
|||||
Db 20149 GATCTGGTGTGCCACGAGGCCACAGACCACCTTCAGCTGGGCATCGCCTGACTGAGAACTC 20208

QY 1397 tgggccttagagcaagcccgaactgtggtgcgcaggcccagggaagcagcactgcatggtgg 1456
|||||
Db 20209 TGGGCTTAGAGCAGGCCCCGAACGTGTGGTCGCCAGGCCCCAGGAGCGGACTGCATGGTGG 20268

QY 1457 aaaaggagctggggcgcatggggaa-caggtccctactatcatcaaccgggagtttggggtcc 1515
|||||
Db 20269 AGAGGGAGCTGGGGCGCATGGGGAAGCAGGTCCCTACTATCAACCCGGGAGTTTGGGGTCC 20328

QY 1516 tccccgaagtaagcaggactgcacgttttcttctcctgattctcgtgttttcttcttg 1575
|||||
Db 20329 TCCCTGAAGTAGGCAGGACTGCACGTTTCTTTCTCTCTGATTCTCGGTTTTTCTTTG 20388

QY 1576 agtcttctggagctgcctctcatcaggtgcactcttcatggaagaagcaactcttgcccc 1635

Db 20389 AGTCTTCTGGAGCTGCCCTTCTCATCAGGTGCACCTCTTCATGG-AAAGCAACTCTTGCCCC 20447
|||||
QY 1636 tacctcttctggcgagggagtaagttaactgctaaaattaaattgtgtgccaggcc 1695
|||||
Db 20448 TACCTCTTCTGGCGCAGCGGAGTAAGTTACTGCTAAATTAATTAATGTGTGCCAGGCC 20507

QY 1696 gggtcggtggctcatgcctgtaatcccc---agcattttgagaggctgagcggttggga 1751
|||||
Db 20508 GGGTCGGTGGCTCATGCTGTAAATCCCCAGGCATTTTGTGAGAGGCTGAGGCGGTGGA 20567

QY 1752 tcacctgaggtcagga-ttcaaaaaccagccctggccaacatagtgaacccccctctctact 1810
|||||
Db 20568 TCACCTGAGGTCAGGAGTTCGAAACCAGCCCTGGCCAACATAGTGAAACCCCTCTCTACT 20627

QY 1811 aaaaatgcaaaaattagtcgcgcgtggtggcacactcctgtaatccccagctacttaggag 1870
|||||
Db 20628 AAAAATGCAAAAATAGTCCGGCGTGGTGGCACACTCCTGTAAATCCCAGCTACTTAGGAG 20687

QY 1871 gctgaggtgggaaaaatcacttggactcca-aagggtggaggttgcagtaagctgaaatcat 1929
|||||
Db 20688 GCTGAGGTGGGAGATCACTTGGACTCCAGGAGGTGGAGGTGCAGTAAGCTGAGATCAT 20747

QY 1930 gccactgcacctagcttgggtggcaaaagcaaaactctatcaaaaaataataataaaat 1989
|||||
Db 20748 GCCACTGCACCCTAGCTTGGTGGCAGAGCAACACTCTATCAAAAAATTAAGTAATAAT 20807

QY 1990 ttgttcaaaaagtcctgc 2006
|||||
Db 20808 TTGTTCAAAAGTCTGTC 20824

RESULT 7
AF176839
LOCUS AF176839 1462 bp DNA linear PRI 22-SEP-1999
DEFINITION Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase
(I-GlcNAc-6-ST) gene, complete cds.
ACCESSION AF176839
VERSION AF176839.1 GI:5917707
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1462)
AUTHORS Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
TITLE Cloning and characterization of a mammalian
N-acetylglucosamine-6-sulfotransferase that is highly restricted to
intestinal tissue
JOURNAL Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)
MEDLINE 99423499
REFERENCE 2 (bases 1 to 1462)
AUTHORS Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
3401 Hillview Avenue, Palo Alto, CA 94304, USA
FEATURES
source
1. .1462
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1"
/tissue_type="intestine"
<229. .>1401
/gene="I-GlcNAc-6-ST"
/product="N-acetylglucosamine 6-O-sulfotransferase"
<229. .>1401
/gene="I-GlcNAc-6-ST"
229. .1401
/gene="I-GlcNAc-6-ST"
/note="possibly involved in biosynthesis of mucosal
ligands for L-selectin"
/codon_start=1

/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD56001.1"
/db_xref="GI:5917708"
/translation="MWLPRESKTVTVLLLAQTTCLILFIISRPGPSPPAGGEDRVHV
LVLSWSRSGSFLGOLFQSFQHPDVPYLMPEPAHWVWTTLSQCSAATLHMAVRDLMRSIFL
CDMDVFDAYMQSNLSAFENWATSRALCSPACSAFPRGTISKQDVCKTLCTRQPF
LAREACRSYSHVLKVEVFNQLVPLSDPALNLRIVHLVRDPRAVLRSREAAGPI
LARDNGIVLTNGKWEADPHLRITREVRSHVIAEAATLKPFPFLRGRYRLVRFD
LAREPLAEIRALYAFTGLTLPQLEAWIHNITHGSGIGKPIEAFTSSRNARNVSQAW
RHALPFTKILRVQEVCAALQLLGYRPVYSADQQRDLTLDLVLPRGPDHFSWASPD"
BASE COUNT 239 a 510 c 446 g 267 t
ORIGIN

Query Match 61.1%; Score 1248.6; DB 9; Length 1462;
Best Local Similarity 99.7%; Pred. No. 2.3e-188;
Matches 1251; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 197 cgggtgccccccagccgcgcagtggtggtgccacggttctccagcaagacagtgacagtg 256
Db 208 CGAAGGCCCCAGCCGCCGCATGTGGCTGCCACGGTCTCCAGCAAGACAGTGACAGTG 267
QY 257 ctccctcctggcacagacacactgcctcctgctctcttcatcatctccccggccaggccctca 316
Db 268 CTCCTCCTGGCACAGACACCTGCCTCCTGCTCTTCATCATCTCCCCGGCCAGGGCCCTCA 327
QY 317 tccccagccggcggcagagatcgtgtgcacgtgctggtgctgtcctcgtgcgtcggtcg 376
Db 328 TCCCCAGCCGGCGGAGATCGTGTGCACGTGCTGTGTCTCTCGTGGCGGCTCGGGC 387
QY 377 tcatcctcttggccagctcttccagccagcaaccccgacgtcttctacctgatggagccc 436
Db 388 TCATCCTTCTTTGGCCAGCTCTTTCAGCCAGCACCCCGACGCTCTTACCTGATGGAGCC 447
QY 437 ggtggtgcatgtgtggaccacactgtcgcaggcgcgcgcgcgcgcgcgcgcgcgcgcgc 496
Db 448 GGTGGCATGTGTGGACCAACCTGTTCGAGGGCAGCGCGGCAACGCTGCACATGGCCGTG 507
QY 497 cgcgacctgatgcgctctatcttttgtgcacatggacgtgttggatgcctacatgcc 556
Db 508 CGGACCTGATCGCTCTATCTTTTGTGCGACATGGACGTTTGTATGCTTACCTGATGCCA 567
QY 557 cagagccgaacactgtccgcctttttcaactgggcaacgagcgcgcgcgcgcgcgcgcgc 616
Db 568 CAGAGCCGAACCTGTCCGCCCTTTTCAACTGGGCAACGAGCGCGCGCTGTGCTCGCCG 627
QY 617 cccgcctgcagcgcctttcccgaggccaccatcagcaagcaggacgtatgaagacactg 676
Db 628 CCGGCTGCAGGCGCTTTCCCGAGGCAACCATCAGCAAGCAGGACGTATGCAAGACACTG 687
QY 677 tgaacgcgccagccattcagcctggcccgaggcgcgcgcgcgcgcgcgcgcgcgcgcgc 736
Db 688 TGCACGGCGCAGCCATTACGCTTGCCCGGGAGGCGCTTACCGGCTGCTCAGCCACGIGTG 747
QY 737 ctcaaggagtgcgctcttcaacctgcaggtgctctacccgctgctcagcgcaccccgcg 796
Db 748 CTCAAGGAGGTGCGCTTCTTCAACCTGCAGGTGCTTACCGGCTGCTCAGCGACCCCGCG 807
QY 797 ctcaacctgcgcacgtgcacctggtgcgcgaccccgcggcgcgcgcgcgcgcgcgcgcgc 856
Db 808 CTCAACCTGCGCATCGTGCACCTGTGTGCGGACCCCGCGGCGCGTGTGCTGCCGGGAG 867
QY 857 gcgcgggcccgatactggcaecgcgacaacggcatcgtgctggggcaccacggcaagtgg 916
Db 868 GCGGCGGCGCGGATCTGGCACGGGACACGGGCATCGTGTGCGGACCAACGGCAAGTGG 927
QY 917 gtggaggccgaccctcacctgcgcctgattcgcgaggtgtgccgcagccacgtgcgcate 976
Db 928 GTGGAGGCGGACCCCTCACCTGGCGCTGATTCGCGAGGTGTGCCCGCAGCCACGTGCGCATC 987
QY 977 gccgaggcgcacactcaagccgcacaccttctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1036
Db 988 GCCGAGGCGCGCACACTCAAGCCGCCACCGCTTCCTGCGCGGCGCGCTACCGCCTGGTGC 1047

QY 1037 ttcgaggacctggcgcgggagccgctggcagagatccgcgcaactctacgccttcaccggc 1096
Db 1048 TTCGAGGACCTGGCGGGGAGCGCTGGCAGAGATCCGCGCACCTACGCCCTTCACCGGC 1107
QY 1097 ctgacctcacgcccacagctcagcgcctggatcccaacaatacaccacacgggtcggggac 1156
Db 1108 CTGACCCCTCACGCCACAGCTCGAGGCCTGGATCCACAACATCACCCACGGGTGCGGGATC 1167
QY 1157 ggcaagccaatcgagcccttcatactctcgtctaggaatggcgcaacgtctccccaggcc 1216
Db 1168 GGCAAGCCAATCGAGGCCTTCCATACTTCGTCTAGGAATGCGCGCAACGCTCTCCAGGCC 1227
QY 1217 tggcgccacgcgttgcccttcaataagatcctgcgcgtgcaggaggtgtgcgcggcgcg 1276
Db 1228 TGCGCCACGCGTTCCTTCACTAAGATCCTGCGCGTGCAGAGGTGTGCGCGGCGCGG 1287
QY 1277 ctgcagctgctgggtaccgctgtgtactctcgtcgggaccagcagcgtgacctcacccctg 1336
Db 1288 CTGCAGCTGCTGGGTACCGGCTGTGTACTCTGCGGACCAGCAGCGTGACCTCACCCCTG 1347
QY 1337 gatcgtgctgccacgagccagaccacttcagctggcgatcgccatcgactgagaaactc 1396
Db 1348 GATCTGCTGTGCCACGAGGCCAGACCACTTCAGTGGGATGCGCTGACTGAGAACTC 1407
QY 1397 tggccttagagcaagccccgaactgtggtgccagggcccgaggaagcagcagcactgc 1451
Db 1408 TGGGCTTAGAGCAGGCGCCGAACTGTGTGTCGCCAGGCCAGGAGCGGACTGCAT 1462

RESULT 8
AF280086
LOCUS
DEFINITION Homo sapiens N-acetylglucosamine 6-O-sulfotransferase GST-4beta
mRNA, complete cds.
ACCESSION AF280086
VERSION AF280086.1 GI:12060803
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; HOMO.
REFERENCE 1 (bases 1 to 3786)
AUTHORS Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and
Rosen,S.D.
TITLE Chromosomal localization and genomic organization for the
galactose/ N-acetylgalactosamine/N-acetylglucosamine
6-O-sulfotransferase gene family
Glycobiology 11 (1), 75-87 (2001)
21096027
JOURNAL
MEDLINE
REFERENCE 2 (bases 1 to 3786)
AUTHORS Hemmerich,S., Bhakta,S., Lee,J.-K., Bistrup,A., Ruddle,N.R. and
Rosen,S.D.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
3401 Hillview Ave., Palo Alto, CA 94304, USA
FEATURES
source
1. 3786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1"
1. 3786
/gene="GST4beta"
/note="clusters with the intestinal N-acetylglucosamine
6-O-sulfotransferase (GST4alpha) gene in GenBank Accession
Numbers AF176838 and AF176839"
189. .1376
/gene="GST4beta"
/note="similar to N-acetylglucosamine 6-O-sulfotransferase
GST-4alpha"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"

QY 261 tcctggcacagaccactgcctcctgctcttcttcacatctccccgagggccctcatccc 320
Db 107538 TCCTGGGCGAG---ACCTTCCTCCTCCTCTTTCTGTTGTTCCCGGCCAGGGCCCTCGTCCC 107594
QY 321 cagccggcggcagagatcgtgtgcacgtgctggtgctgtcctcgtgctcgtcgggctcctcat 380
Db 107595 CAGCAGGCGGCGAGGCGCGGTCATGTGCTGCTGCTTCCTCCTGCTGCGCTCGGGCTCGGT 107654
QY 381 ccttcttggccagctcttcagccagcacccccgacgtcttcttacctgatggagccccggt 440
Db 107655 CCTTCGTGGCCAACTCTTCAACACGACACCCCGACGCTCTTACCTAATGGAGCCCGCGT 107714
QY 441 ggcattgtggaccacccctgtgcagggcagcgcggcgaacgctgcacatggccgtgcgg 500
Db 107715 GGCACGTGTGGACCAACCCCTGCGAGGGCAGCGCCGCAACGCTGCACATGGCTGTGCGCG 107774
QY 501 acctgatgcctctatctttttgtgcacatggacgtgtttgatgcctacatgccacaga 560
Db 107775 ACCTGGTGGCTCCGCTCTTCTGTGGACATGGACGTGTTGATGCCATCTGCCCTTGGC 107834
QY 561 gccgaacctgtccgcctttttcaactgggcaacgagccgcgctgtgtcctgcgcgcccg 620
Db 107835 GCCGCAACCTGTCCGACCTTTCCAGTGGGCGCGTGAGCCCGTGCACCTGTGCTCGCCACCCG 107894
QY 621 cctgcagcgcctttccccagggcaaccatcagcaagcaggacgtatgcaagacactgtgca 680
Db 107895 CTTGCAGTGCCTTTTCCCGGAGGCGGCATCAGCAGCAGGAGGCGGTGTGCAAGCCACTGTGCG 107954
QY 681 cgcgcagccattcagcctggccgggagccctgcctcctcactacagccactgtgtctca 740
Db 107955 CGCGGCAGTCCCTTCAACCTTGGCGGGAGGCGCTGCCGCTCCTACAGCCACGTGTGTCTCA 108014
QY 741 agcagtgccgtcttccaacctgcaggtgctctacccgctgcctcagcagcccccgctca 800
Db 108015 AGGAGTGGCTTTCTTCAACCTGCAGGTGCTCTACCCGCTGCTCAGCGACCCCGGCTCA 108074
QY 801 acctgcgcctcgtgcacctggtgcgcagccccgcggccgctgcctcctcccgaggcgg 860
Db 108075 ACCTAGCATCGTGCACCTGGTGGCGACCCCGCGGCGCTGCTGCGCTCCCGGGAGCAGA 108134
QY 861 cggcccgcatactggcacgcgacaaaggcatcgtgctgggcaccaacggcgaagtgggtgg 920
Db 108135 CAGCCCAAGGCTCTGGCGCGTGACACAGGCATCGTGTGGGCGCACCAACGGCACGTGGGTGG 108194
QY 921 agccgcacctcactgcgcctgattgcgcaggtgtgcgcagccacagtcacatgcgcg 980
Db 108195 AGGCCGACCCCGGCTGCGCGTGGTGGCGGAGGTGTGCCGTAGCCACGTACGCATCGCCG 108254
QY 981 agccgcacactcaagccgcaccccttctcgtgcgcgcgcctacccgctggtgcgttcg 1040
Db 108255 AGCGCGCACACTCAAGCCGCCACCCCTTCTGCGCGCGCGCTACCGCTGCTGCGCTTTCG 108314
QY 1041 agcacctggcgcgggagccgctggcagagatccgcgcactctacgcttcacccggcctga 1100
Db 108315 AGGACCTGGCGGGGAGCCGCTGGCAGAAATCCGTGCGCTCTACGCTTCACTGGGCTCA 108374
QY 1101 ccctcacgccaacagctcagggcctggatccacacatcacccacaggggtgggatcggca 1160
Db 108375 GTCTACGCCACAGCTCGAGGCCCTGGATCCATAACATCACCCACGGATCTGGACCTGGTG 108434
QY 1161 agccaatcagggccttccatacttcgtctaggaatgcgcgcgaacgtctctccagggcctggc 1220
Db 108435 CGCGCCGGGAAGCCTTCAAGACTTCGTTCAGGAATGCGCTCAACGCTCTCCAGGCCCTGGC 108494
QY 1221 gccacggttgcccttcactaagatcctgcgcggtgcaggaggtgtgcgcgcgcgcgtgc 1280
Db 108495 GCCATGCGCTGCCCTTTGCCAAGATCCGCGCGGTGCAGGAACCTGTGCGCTGGTGGCGCTGC 108554
QY 1281 agctgtgggtaccggcctgtgtactctgcgtggaccagcagcgtgaacctcaacctggatc 1340
Db 108555 AGCTGTGGGCTACCGGCCCTGTGTACTCTGAGGACGAGCAGCGCAACCTCGCCCTTGATC 108614
QY 1341 tgggtgctgccacgagggccagaccacttcagctgggcatac--gcctgactgagaactctg 1398

Db 108615 TGGTGTGTCCACAGAGGCTTGAAACGGCTTCACCTTGGGCATCATCCACGCTTCGCACCCCC 108674
QY 1399 ggccttagagcaagccccgaactgtg--gtcgcagggccccaggaagcagcactgcattgtgg 1456
Db 108675 GAAATTAGTGAGGCCACAGTTGTAGCAGGCGCTAGGCCCGGGAGGAGAGTGCATGGTGC 108734
QY 1457 aaaagagctgggcgcatgggga--caggtccctactatcaacggggagtttggggt-- 1513
Db 108735 AGAGGGGCTGGGCGCACGGAGAGAGGAGTCCCTATATTGACCAAGGAGTTTGTGGTAC 108794
QY 1514 -----cctccccctgaagtaagaaggactgcacgtttcttctctcctgattctcgg 1565
Db 108795 GACCCCTCCCCCTCCCCAAGTAGGCAAGGACTGCACGTTCTTCTCTCTCTGATTCTTGG 108854
QY 1566 ttttcccttgagtcttcttgagctgccttctcctcactcaggtgcactcttcaatgaaaagcaa 1625
Db 108855 TTTTCCTTTGAGTCCCTCTGGAGCTGCCCTCTCATCAGGTGCACCTTCATGG-AAAGCAA 108913
QY 1626 ctcttgccccctacctcttcttgggcgagggagtaagtt 1663
Db 108914 CTCTTGCCCT-CCTCCTCTGGGCACAGGGTGTGCGTT 108950

RESULT 12
AX327330
LOCUS AX327330 1647 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 15 from Patent WO0179468.
ACCESSION AX327330
VERSION AX327330.1 GI:18097876
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Policky,J.L., Hafalia,A., Burford,N., Ring,H.Z., Lal,P.,
Tribouley,C.M., Yao,M.G., Yue,H., Tang,Y.F., Patterson,C., Das,D.,
Sanjanwala,M.S., Gandhi,A.R., Reddy,R., Khan,F.A., Baughn,M.R.,
Ramkumar,J., Griffin,J.A. and Au-Young,J.
TITLE Drug metabolizing enzymes
JOURNAL Patent: WO 0179468-A 15 25-OCT-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source 1..1647
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 7472777CB1"

BASE COUNT 287 a 560 c 501 g 299 t
ORIGIN

Query Match 46.3%; Score 947; DB 6; Length 1647;
Best Local Similarity 85.1%; Pred. No. 1.1e-140;
Matches 1106; Conservative 0; Mismatches 185; Indels 8; Gaps 4;

QY 218 atgtggctgccacggttctccagcaagacagtgacagtgtcctcctcctggcacaccacc 277
Db 1 ATGTGGCTGCCGGCGGTCTCCAGCACAGCAGTACCGCGCTCCTCCTGGCGCAG---ACC 57
QY 278 tgcctcctgctcttcacatctccccggccagggccctcatccccagccgcgcgaggtat 337
Db 58 TTCCCTCCTCCTCTTCTTGGTTTCCCGGCCAGGGCCCTCGTCCCCAGCAGGCGGCGAGGCG 117
QY 338 cgtgtgcacgtgctgggtgctgtcctcgtggcgtcggggtcatccttcttgggcccagctc 397
Db 118 CGCGTGCATGTGTGTGTGTCTCCTCGTGGCGCTCGGGCTCGTCTCCTCGTGGGCCAACTC 177
QY 398 ttcagccagcaccocgacgtcttctacotgatggagcccgctggcatgtgtgagaccacc 457
Db 178 TTCAACCAAGCACCCCGACGCTCTTCTACCTAATGGAGCCCCGCGTGGCACGTTGGACCA 237
QY 458 ctgtcgcagggcgagcgggcaacgctgcacatggcgcgtgcgcacgtgatgcgtctatc 517


```
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCYW
Center clone name: CH230-8L17
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 96308 bases at least Q40
Consensus quality: 106296 bases at least Q30
Consensus quality: 113927 bases at least Q20
Estimated insert size: 91054; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 71 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 6038: contig of 6038 bp in length
* 6039 6138: gap of unknown length
* 6139 9960: contig of 3822 bp in length
* 9961 10060: gap of unknown length
* 10061 13888: contig of 3828 bp in length
* 13889 13988: gap of unknown length
* 13989 18180: contig of 4192 bp in length
* 18181 18280: gap of unknown length
* 18281 21330: contig of 3050 bp in length
* 21331 21430: gap of unknown length
* 21431 23489: contig of 2059 bp in length
* 23490 23589: gap of unknown length
* 23590 27447: contig of 3858 bp in length
* 27448 27547: gap of unknown length
* 27548 30005: contig of 2458 bp in length
* 30006 30105: gap of unknown length
* 30106 33183: contig of 3078 bp in length
* 33184 33283: gap of unknown length
* 33284 36345: contig of 3062 bp in length
* 36346 36445: gap of unknown length
* 36446 38892: contig of 2447 bp in length
* 38893 38992: gap of unknown length
* 38993 40983: contig of 1991 bp in length
* 40984 41083: gap of unknown length
* 41084 43882: contig of 2799 bp in length
* 43883 43982: gap of unknown length
* 43983 47307: gap of unknown length
* 47407 49142: contig of 1736 bp in length
* 49143 49242: gap of unknown length
* 49243 52210: contig of 2968 bp in length
* 52211 52310: gap of unknown length
* 52311 55011: contig of 2701 bp in length
* 55012 55111: gap of unknown length
* 55112 57128: contig of 2017 bp in length
* 57129 59164: gap of unknown length
* 59165 59264: gap of unknown length
* 59265 60945: contig of 1681 bp in length
* 60946 61045: gap of unknown length
* 61046 62362: contig of 1317 bp in length
* 62363 62462: gap of unknown length
* 62463 64180: contig of 1718 bp in length
* 64181 64280: gap of unknown length
* 64281 65664: contig of 1384 bp in length
* 65665 65764: gap of unknown length
* 65765 67613: contig of 1849 bp in length
* 67614 67713: gap of unknown length
```

```

67714 69378: contig of 1665 bp in length
69379 69478: gap of unknown length
69479 71051: contig of 1573 bp in length
71052 71151: gap of unknown length
71152 73343: contig of 2192 bp in length
73344 73443: gap of unknown length
73444 75183: contig of 1740 bp in length
75184 75283: gap of unknown length
75284 76990: contig of 1707 bp in length
76991 77090: gap of unknown length
77091 78115: contig of 1025 bp in length
78116 78215: gap of unknown length
78216 80432: contig of 2217 bp in length
80433 80532: gap of unknown length
80533 81894: contig of 1362 bp in length
81895 81994: gap of unknown length
81995 83219: contig of 1225 bp in length
83220 83319: gap of unknown length
83320 84714: contig of 1395 bp in length
84715 84814: gap of unknown length
84815 86250: contig of 1436 bp in length
86251 86350: gap of unknown length
86351 88174: contig of 1824 bp in length
88175 88274: gap of unknown length
88275 89924: contig of 1650 bp in length
89925 90024: gap of unknown length
90025 91574: contig of 1550 bp in length
91575 91674: gap of unknown length
91675 93173: contig of 1499 bp in length
93174 93273: gap of unknown length
93274 94320: contig of 1047 bp in length
94321 94420: gap of unknown length
94421 95911: contig of 1491 bp in length
95912 96011: gap of unknown length
96012 97119: contig of 1108 bp in length
97120 97219: gap of unknown length
97220 98221: contig of 1002 bp in length
98222 98321: gap of unknown length
98322 99324: contig of 1003 bp in length
99325 99424: gap of unknown length
99425 100879: contig of 1455 bp in length
100880 100979: gap of unknown length
100980 102554: contig of 1575 bp in length
102555 102654: gap of unknown length
102655 104910: contig of 2256 bp in length
104911 105010: gap of unknown length
105011 107042: contig of 2032 bp in length
107043 107142: gap of unknown length
107143 108339: contig of 1197 bp in length
108340 108439: gap of unknown length
108440 109717: contig of 1278 bp in length
109718 109817: gap of unknown length
109818 111029: contig of 1212 bp in length
111030 111129: gap of unknown length
111130 112643: contig of 1514 bp in length
112644 112743: gap of unknown length
112744 114325: contig of 1582 bp in length
114326 114425: gap of unknown length
114426 115714: contig of 1289 bp in length
115715 115814: gap of unknown length
115815 116991: contig of 1177 bp in length
116992 117091: gap of unknown length
117092 118249: contig of 1158 bp in length
118250 118349: gap of unknown length

```

```

Query Match 35.2%; Score 719.6; DB 2; Length 139252;
Best Local Similarity 76.5%; Pred. No. le-104;
Matches 922; Conservative 0; Mismatches 279; Indels 5; Gaps 3;
QY 178 ctctcccagcagccctggacggtggccccccagccgcccgcacgtgtgctgccaggttctc 237
Db 57932 CTACCCCTGCCTCTCTCTTTTCATAGGCTCTAGCCAGAGGCATGCGGCTACCCCGCTTCTC 57991
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 09:23:44 ; Search time 10470.7 Seconds
(without alignments)
2634.765 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggctcgagggtccactgtgct.....aaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	657.4	32.2	695	10	BE857538	BE857538 7g01a08.x
C 2	631	30.9	735	10	BE858652	BE858652 7g01a09.x
C 3	525	25.7	525	10	BF197521	BF197521 7o84a08.x
C 4	487.6	23.9	620	9	AI824100	AI824100 wj45c01.x
C 5	420.8	20.6	436	9	AW081348	AW081348 xc41b06.x
C 6	352.4	17.2	1923	11	AK009113	AK009113 Mus muscu
C 7	346.4	16.9	1067	12	CNS03KN7	AL248380 Tetraodon
C 8	322.8	15.8	657	12	AG035205	AG035205 Pan trogl
C 9	308.4	15.1	517	10	BF042384	BF042384 BP250022A
C 10	295.4	14.5	568	9	AI115260	AI115260 ui43c07.y
C 11	284.2	13.9	954	10	BI823850	BI823850 603039012
C 12	279.4	13.7	849	12	CNS04QFN	AL302540 Tetraodon
C 13	260	12.7	260	10	BI824736	BI824736 603033615
C 14	259.4	12.7	955	10	BG107354	BG107354 602290740
C 15	245.4	12.0	433	12	AZ405100	AZ405100 IM0173108
C 16	244.6	12.0	429	9	BB849113	BB849113 BB849113
C 17	231.4	11.3	337	10	BG960153	BG960153 FM3-CT064

18	224.4	11.0	695	10	BG964671	BG964671 602831875
19	224.4	11.0	783	10	BG963298	BG963298 602827716
20	218.8	10.7	852	10	BG966340	BG966340 602832826
21	217.4	10.6	624	12	AQ373399	AQ373399 RPCI11-15
22	214.2	10.5	1655	11	BC022315	BC022315 Homo sapi
C 23	213.8	10.5	560	9	AL589348	AL589348 DKFZp451K
24	211.4	10.3	452	10	BF725761	BF725761 bx19e11.y
25	211	10.3	352	9	AW089016	AW089016 xd34d02.x
26	210.4	10.3	515	10	BM467754	BM467754 AGENCOURT
27	210.2	10.3	400	9	AV760391	AV760391 AV760391
28	210	10.3	388	12	AQ557029	AQ557029 HS_5308_B
C 29	209.4	10.2	693	12	AG118065	AG118065 Pan trogl
30	209.2	10.2	412	9	BE062478	BE062478 QV4-BT025
31	209	10.2	539	12	AQ379787	AQ379787 RPCI11-15
32	208.4	10.2	1172	11	AF305824	AF305824 Homo sapi
C 33	208.2	10.2	416	9	AW341978	AW341978 hb74d10.x
34	208	10.2	2048	11	BC017051	BC017051 Homo sapi
35	207.8	10.2	616	12	AQ554309	AQ554309 RPCI-11-4
C 36	207.2	10.1	451	9	AV695478	AV695478 AV695478
C 37	207.2	10.1	650	12	AQ530872	AQ530872 RPCI-11-3
C 38	207	10.1	384	9	AL040054	AL040054 DKFZp434P
39	207	10.1	946	10	BG335756	BG335756 602404452
40	206.6	10.1	602	12	AQ540860	AQ540860 RPCI-11-3
C 41	206.4	10.1	345	10	BG222875	BG222875 naf60e11.
42	206	10.1	544	10	BG739841	BG739841 602630570
43	206	10.1	544	12	AQ394650	AQ394650 CITBI-E1-
C 44	205.8	10.1	688	12	AG118999	AG118999 Pan trogl
45	205.6	10.1	440	10	BG189911	BG189911 RST8965 A

ALIGNMENTS

RESULT 1
BE857538/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE857538 695 bp mRNA linear EST 29-SEP-2000
7g01a08.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305174 3'
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGUCOSAMINE
6-O-SULFOTRANSFERASE. ; mRNA sequence.

BE857538

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 695)

NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project

(CGAP/BTGP), Tumor Gene Index

Unpublished (1998)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,

Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 431.

Location/Qualifiers

1. .695

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3305174"

/clone_lib="NCI_CGAP_Brn23"

/tissue_type="glioblastoma (pooled)"

/lab_host="DH10B"

/Note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGGCATATCTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 127 a 210 c 247 g 111 t
ORIGIN

Query Match 32.2%; Score 657.4; DB 10; Length 695;
Best Local Similarity 97.6%; Pred. No. le-67;
Matches 678; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 284 ctgtctttcatcatctcccgccagggccctcatcccccagcgccgagggatcggtg 343
Db 695 CTGCTCTTTCATCATCTGCCGGCCAGGGCCCATCATCACCAGCGCGCAGCGATCGTGTG 636
QY 344 cacgtgctggtgctgtcctcgtggcgctcgggctcattctttggccagctcttcagc 403
Db 635 CACGTGCTGATGCTGTCTCGTGGCGCTCGGGCTCATCC-TCTTGGACAGCTCTTCAGC 577
QY 404 cagcaccccgacgtctttctacctgatggagcccgctggcgatgctgtggaccaccctgtcg 463
Db 576 CAGCACACCGACAGCTCTTCTACTTGATGGAGCCCGCTGGCAATGTGGACCATCCTGTGCG 517
QY 464 cagggcagcgcggcaacgctgcacatggccgtgcgcgacctgatgcgtctatctttttg 523
Db 516 CAGGGCAGCGCGGCAACGCTGCACATGGCCGTGCGCGACGCTGATGGCTCTATCTTTTG 457
QY 524 tgcgacatggacgtgtttgatgcctacatgccacagagccgaacctgtccgccttttc 583
Db 456 TGGACATGCACGCTGTCTGATGCCCTACATGCCACAGAGCCGAAACCTGTCCGCCTATTTC 397
QY 584 aactgggcaacgagccgcgctgtctgcgcgcgcgcctgcagcgctttcccccagggc 643
Db 396 AACTGGGCAACGAGCCGCGGCTGTGTCGCGCCCGCTGCAGCGCCTTCCCGAGGC 337
QY 644 accatcagcaagcaggacgtatgcaagacactgtgcacgcgccagccaattcagccctggcc 703
Db 336 ACCATCAGCAAGCAGGACGTATGCAAGACACTGTGCACGGCGCAGCCATTACGCTGGCC 277
QY 704 cgggaggccctgcgctcctacagccacgtgggtgctcaaggaggtgcgcttcttcaacctg 763
Db 276 CGGGAGGCCCTGCGCTCCTACAGCCACGTGGTGCTCAAGGAGGTGCGCTTCTTCAACCTG 217
QY 764 caggtgctctaccgcgtgctcagcgaccccgcgctcaacctgcgcacgtgcacctgggtg 823
Db 216 CAGGTGCTCTACCCGCTGCTCAGGACCCCGCGCTCAACCTGCGCATCGTGCACCTGGTG 157
QY 824 cgcgaccccgcgccgctgctgcgtcccgaggagcgcgcccgatatactaggcacgcgac 883
Db 156 CGCGACCCCGCGCGCTGCTGCGCTCCCGGAGCGCGCGGCGCGGATACITGGCACGCGAC 97
QY 884 aacggcatcgtgtgggcaaccaacggcaagtgggtggagggccgacctcactgcgcctg 943
Db 96 AACGGCATCGTGTGGGCACCAACGGCAAGTGGTGGAGGCCGACCCCTCACCTGCGCCTG 37
QY 944 attcgcgaggtgtgcgcgagccacgtgcgcacgcg 978
Db 36 ATTCCGAGGTTGTCCCGCAGCCACGTCGGCATCGC 2

RESULT 2
BE858652/c
LOCUS BE858652 735 bp mRNA linear EST 29-SEP-2000
DEFINITION 7g01a09.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305176 3'
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
6-O-SULFOTRANSFERASE. ;, mRNA sequence.
ACCESSION BE858652

VERSION BE858652.1 GI:10373890
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 735)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 425.
FEATURES
Location/Qualifiers
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGGCATATCTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 135 a 221 c 253 g 120 t
ORIGIN

Query Match 30.9%; Score 631; DB 10; Length 735;
Best Local Similarity 93.5%; Pred. No. 1.2e-64;
Matches 677; Conservative 0; Mismatches 45; Indels 2; Gaps 2;

QY 300 cccggccaggccctcatcccccagcgccgagggatcgtgtgcacgtgctgtgctgt 359
Db 734 CCGGCNCAGGGCCTCATCCACAAGCAGNCGCGAAGATCGTGTGCACGTGTGGTGTGT 675
QY 360 cctcgtggcgtcgggctcatccttcttggccagctcttcagccagcaccgcgacctct 419
Db 674 CCTCGTGGCGCTCGGGCTCATCTCTNAGGGCCAGCTCTAGCCCCAGCA-CCCCACGTCT 616
QY 420 tctacctgatggagcccgctggcactgtgtgtggaccacctgtcgcagggcagcgggcaa 479
Db 615 TCTACATGATGAAGCCCGCGTGGCATGTGTGG-CCACCTGTTCGAGGGCAGCGCGGCAA 557
QY 480 cgctgcacatggccgtgcggacctgatgcgctctatcttttgcgacatggacgtgt 539
Db 556 CGCTGCACATGGCCGTGCGGACCTGATCGCTCTATCTTTTGTGCGACATGGACGTGT 497
QY 540 ttgatgcctacatgccacagagccgaacctgtccgccttttcaactgggcaacgagcc 599
Db 496 TTGATGCNTACATGCCACAGAGCCGAAACCTGTCCGCCCTTTTCAACTGGGCAACGAGCC 437
QY 600 ggcgcgtgtgctgcgcgcgcctgcagcgcctttcccccagggcaccatcagcaagcagg 659
Db 436 GCGCGCTGTGCTCGCGCGCCCGCCTGCAGCGCCTTTTCCCCGAGGSCACCATCAGCAAGCAGG 377

QY 660 acgtatgcaagacactgtgcaagcggcagccattcagcctgcccggaggcctgccgct 719
|||||
Db 376 acgtatgcaagacactgtgcaagcggcagccattcagcctgcccggaggcctgccgct 317
|||||
QY 720 cctacagccacgtgtgctcaagggaggtgcgcttcttcaacctgcaggtgctctaccgc 779
|||||
Db 316 cctacagccacgtgtgctcaagggaggtgcgcttcttcaacctgcaggtgctctaccgc 257
|||||
QY 780 tgctcagcgaccccgctcaaacctgcgcacgtgcacctggtgcgcgaccccgggcgcg 839
|||||
Db 256 tgctcagcgaccccgctcaaacctgcgcacgtgcacctggtgcgcgaccccgggcgcg 197
|||||
QY 840 tgctgcgtcccgaggaggcggcccgatactggcagcgcgacaaagcgcacgtgctgg 899
|||||
Db 196 tgctgcgtcccgaggaggcggcccgatactggcagcgcgacaaagcgcacgtgctgg 137
|||||
QY 900 gcaccaacggcaagtggaggcggcccgatactggcagcgcgacaaagcgcacgtgctgg 959
|||||
Db 136 gcaccaacggcaagtggaggcggcccgatactggcagcgcgacaaagcgcacgtgctgg 77
|||||
QY 960 gcagccacgtgcgcacgtgcgcagcggcccgacacactaaagcggccacccctcctgcgggcc 1019
|||||
Db 76 gcaaccacgtgcgggattgccgaagcccaacacactcaaacccgccacccttttggsggggc 17
|||||
QY 1020 gcta 1023
| |
Db 16 gnga 13

RESULT 3
BF197521/c
LOCUS
DEFINITION
7084a08.x1 NCI_CGAP_Kid11 Homo sapiens cdna clone IMAGE:3642903 3',
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGUCOSAMINE
6-O-SULFOTRANSFERASE. ;, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BF197521
BF197521.1 GI:11086670
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 525)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
High quality sequence stop: 451.
Location/Qualifiers
1. 525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3642903"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library

FEATURES
source
1. 525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3642903"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library

(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo. "
BASE COUNT 90 a 163 c 187 g 85 t
ORIGIN
Query Match 25.7%; Score 525; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 2.7e-52;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 443 catgtgtggaccaccctgtgcagggcagcggcaacgctgcaaatgcccgcgac 502
|||||
Db 525 CATGTGTGGACCACCCCTGTGCGAGGCGAGCGGCAACGCTGCACATGCCGTGCGCGAC 466
|||||
QY 503 ctgatgcgtctatctttttgtgcacatggacgtgtttgtatgcctacatgccacagac 562
|||||
Db 465 CTGATGCGCTCTATCTTTTGTGCGACATGGACGCTGTTTGATGCTACATGCCACAGAGC 406
|||||
QY 563 cgaacctgtccgcctttttcaactgggcaacgagcgcgctgtgctcgcgcgcgc 622
|||||
Db 405 CGAAACCTGTCCGCCTTTTCAACTGGGCAACGAGCGCGCGCTGTGCTGCGCGCGCC 346
|||||
QY 623 tgcagcgcctttccccgaggcaccatcagcaagcaggacgtatgaagacactgtgcacg 682
|||||
Db 345 TGCAGCGCCTTTCCCGAGGCGACCATCAGCAAGCAGGACGTATGCAAGACACTGTGCACG 286
|||||
QY 683 cggcagccattcagccctggccggaggcctgcctacccgctgcctcagcagcccgctcaag 742
|||||
Db 285 CGGCAGCCATTTCAGCCCTGGCCGGAGGCCTGCCGCTCTACAGCCACGTGCTCAAG 226
|||||
QY 743 gaggtgcgcttttcaacctgcaggtgctctacccgctgcctcagcagcccgctcaag 802
|||||
Db 225 GAGGTGCGCTTCTTCACTGCAGGTGCTCTACCCGCTGCTCAGCGACCCCGCGCTCAAC 166
|||||
QY 803 ctgcgcacgtgcacctggtgcgcgaccccgggcgctgctgcctcccgaggcggcg 862
|||||
Db 165 CTGCGCATCGTGCACCTGGTGCAGCGACCCCGGGCGGTGCTGCGCTCCCGGGAGCGCGG 106
|||||
QY 863 ggcccgatactggcagcgcgacaaacggcctgctgtggcaccacggcaagtgggtggag 922
|||||
Db 105 GGCCCGATACTGGCAGCGGACAAACGGCATCGTGTGGGCACCAACGGCAAGTGGGTGGAG 46
|||||
QY 923 gccgacccctcactgcgcctgattcgcgaggtggtgcccgcagccac 967
|||||
Db 45 GCGGACCCCTCACCTGGCCTGATTTCGCGAGGTGTGCGCGACGCCAC 1

RESULT 4
AI824100/c
LOCUS
DEFINITION
AI824100
620 bp mRNA linear EST 21-DEC-1999
WJ46C01.x1 NCI_CGAP_Lu19 Homo sapiens cdna clone IMAGE:2405856 3',
similar to TR:O75667 O75667 DJ71L16.4 ; contains PTR5.b2 PTR5
repetitive element ;, mRNA sequence.
AI824100
AI824100.1 GI:5444771
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 620)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Db	279	GGCAGACACCCGGATGTGTTCTACCTGATGGAGCCCTGCTGGCATGTGTGGATGACTTTC	338			
QY	461	tgcagggcagcgcggcaacgctgcacatggccgtgcgacacctgatgcgtcttatcttt	520			
Db	339	ACCAGCAGACAGCCTGGAAGCTGCACATGGCTGTGCGGATCTTCTGCGTTCOGTCTTC	398			
QY	521	tttgtgcacatggacgtgtttgatgcctacatg---ccacagagccgaacacctgtccgcc	577			
Db	399	CTGTGTACATGAGCGTCTTTGATGCCTACATGAACCCAGGCCCCCCGGAACAGTCCAGC	458			
QY	578	tttttcaactgggcaacgagccgcgcgtgtgtctgcgcgcgcgcctgcagcgcctttccc	637			
Db	459	CTCTTCAGTGGGAGCAAAAGCCGGGCCCTGTGCTCAGCGCCTGTGTGTGACTTCTTCCCT	518			
QY	638	cgaggcaccatcagcaagcaggacgtatgcaagacacactgtgcacgcggcagccattcagc	697			
Db	519	GCCACGAGATCAGCTCACCCAAGCACTGCAAGCTGCTCTGCGGTACAGCAGCCCTTTGAT	578			
QY	698	ctggccgggaggcctgcgcctcctacagccacagtggtgtctcaaggaggtgcgtcttctt	757			
Db	579	ATGGTGGAGAAGGCCCTGCCGCTCTCACGGCTTCGTGGTACTCAAGGAGGTGCGTTTCTC	638			
QY	758	aacctgcagtgctctacccgctgctcagcagcccccgcgcctcaacctgcgcacgtgcac	817			
Db	639	AGCCTGCAGGCCCTCTATCCACTACTACGGACCCCTTCCCTCAACCTGCACGTCGTGCAC	698			
QY	818	ctggtgcgcagccccgggcccgtgctgcgctcccgaggcggcggccccgatactggca	877			
Db	699	CTGGTCCGAGACCCCCGGCCGCTGTTCGGATCCCGGAGCACACCCACATAGAACTCATG	758			
QY	878	cgcgacaacggcatcgtctggtggcaccacggcaagtgtgtgaggccgacccctcacctg	937			
Db	759	GTTGACAGTCATATTGTGCTAGGGCAGCATTTGGAACGATCAAGGAGGAAGACCAGCCC	818			
QY	938	cgcctgattgcgagggtgtgcgcgagccacgtgcgcctgcgcaggccgacacactcaag	997			
Db	819	TATTATGCCATGAAGATCATCTGCAAAAGCCAGGTGGACATAGTCAAGGCCATCCAAACC	878			
QY	998	ccgccacccttctcgcgcgcgcgtaccgccttggtgcgcttcgaggaacctggcgcgggag	1057			
Db	879	CTCCCTGAAGCTCTGCAGCAGCGCTACCTGTTCCTGAGGTATGAGGACCTGGTTCGGGCA	938			
QY	1058	ccgctggcagagatccgcgcacactctacgccttcaaccggcctgacccctcaagccacagctc	1117			
Db	939	CCCCTGCCCCAGACGACACAGACTAATAAATTGTGGGTTGGATTTTTTGCCCCACCTC	998			
QY	1118	gaggcctggatccacaacatcacccacgggtcggggatcggaagccaatcgaggcccttc	1177			
Db	999	CAACATGGGTTTACAATGTCACCCCGCGCAAGGCATGGTCAGC-----ATGCTTTC	1052			
QY	1178	catacttcgttaggaatgcgcgaacgtctcccaggcctggcgcacgcgttgcccttc	1237			
Db	1053	CATACTAACGCCAGGAACGCCCTCAACGCTCTCTCAGGCGTGGCGTTGGTCCCTTACCTTAC	1112			
QY	1238	actaagatcctgcgcgtgcaggaggtgtgcgcgcggcgcgcgtgcagctgctgggtaccgg	1297			
Db	1113	GAAAAGGTTTCCCAGCTTCAAGATGCCTGCGGTTGAGGCTATGGATTTGCTGGGATACCTC	1172			
QY	1298	cctgtgtactctgcggaccagcagcgtgacctcaaccctggatctggtg	1345			
Db	1173	CAGGTCAGATCTCAACAAGAACAAAGGCAACCTGTCCCTGGATCTTCTG	1220			
RESULT	7	CNS03KN7	1067 bp	DNA	linear	GSS 17-MAY-2000
CNS03KN7/c						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						

ORGANISM	Tetraodon nigroviridis					
REFERENCE	1 (bases 1 to 1067)					
AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.					
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetraodon nigroviridis					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 1067)					
AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.					
TITLE	Human gene number estimate provided by genome wide analysis using tetraodon nigroviridis DNA sequence					
JOURNAL	Unpublished					
REFERENCE	3 (bases 1 to 1067)					
AUTHORS	Genoscope.					
TITLE	Direct Submission					
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases					
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.					
FEATURES	Location/Qualifiers					
source	1..1067					
	/organism="Tetraodon nigroviridis"					
	/db_xref="taxon:99883"					
	/clone="033J20"					
	/clone_lib="G"					
	/note="Genoscope sequence ID : C0BG033DE10LP1-end : T7"					
BASE COUNT	241	a	306	c	323	g
ORIGIN	197	t				
Query Match	16.9%;	Score	346.4;	DB 12;	Length	1067;
Best Local Similarity	62.1%;	Pred.	No. 1.1e-31;			
Matches	564;	Conservative	0;	Mismatches	341;	Indels
					3;	Gaps
						1;
QY	234	tctccagcaagacagtgacagtgctcctcctggtgcacagaccacctgcctcctgtcttca	293			
Db	906	TCGGCACCATGATTTTGTGTTGACCCCTCCAGGGAGCCACGGTGATGCTGCTGAGCGGCT	847			
QY	294	tcattctccgggccaggccctcatcctcccccagccggcggcaggatcgtgtgcactgtctgg	353			
Db	846	GGTACTTCCAGCTCAGCCCTGCTCTCCGCTCCCTCCGCGGGAAAGTTTCACGTTCTCC	787			
QY	354	tgctgtcctcgtggcgtcgggtcctcctcctcctcctcctcctcctcctcctcctcctcctc	413			
Db	786	TGCTGTCTGCTCCTGGCGCTCGGGCTCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	727			
QY	414	acgtctctacctgatggagcccgctggcgtggtgcacacctgatgcgtctctctctctctctc	473			
Db	726	CGGTCTTCTATCTGATGGAGCCCGGGTGGCACGCTGTGGAGCCAGCTCCCCAAGGTGAACG	667			
QY	474	cggaacgctgcacatggccgtgcgcgacctgatgcgtctctctctctctctctctctctctc	533			
Db	666	TTGGGATGCTCCGGATTGCTGTGAGGGATATATTTCGGAGCATATTCAGTGGGACTTCT	607			
QY	534	acgtgtttgatgcctacatgccacagagccgaacacctgtccgccttttcaactgggcaa	593			
Db	606	CAGGAATGGAGGCCTTCCTGCTGCTGAAAATCACACATATCGGACGTTTCATGTGGAGCC	547			
QY	594	cgagccgcgcgtgtgctgc	653			
Db	546	ACAGCAGGGCGCTGTGCTCGCCCCCGCGCTGCTCTCTCAACCCCGCGGGAAGGATGAGCG	487			
QY	654	agcaggacgtatgcaagacactgtgcacgcggcagccattcagcctggcccgaggagcct	713			
Db	486	ATCAGCCCGAGTGTCTTGAAAAAATGCGTCTCCAAGGGTCTGCACGGGGCTAAGAACGCAT	427			


```
Db 244 CTTCGTGGGCAGCTCTTCAGCCCAACACCCCGATGTCTTCTACCTGATGGAGCCGGCITG 303
QY 442 gcatgtgtgaccaccctgtgcagggcagcgcggaacgctgcacatggccgtgcgga 501
Db 304 GCACGTCTGGGATACGTTGTTCAGGGCAGTGCCCCGCACTCCACATGCCGTGCGTGA 363
QY 502 cctgatgcgtctatctttttgtgcacatggacgtgtttgatgcctacatgccacag 561
Db 364 CCTGATCCGCTCAGTGTCTTCCATGACATGGACGTAATTGATGCCCTACCTGCCCTGGCG 423
QY 562 ccgaacacgtctccgcctttttcaactgggcaacgagccgcgcgctgtgctgcgcgcgc 621
Db 424 CCGCAACATCTCGGATCTCTTCCAGTGGCGGTGAGCCGCGCAITGTGCTCACCTCCGGT 483
QY 622 ctgcagcgcctttcccccaggccaccatcagcaagcaggacgtatgcaagacactgtgcac 681
Db 484 CTGCGAAGCCTTCGCTCGTGGCAACATCAGCAGCGGAGGAGGTGTGAAGCCTCTGTGCGC 543
QY 682 gcggcagccattcagcctggcccg 706
Db 544 AACCGGCCCTTCGGCCTGGCTCAG 568
```

```
RESULT 11
BI823850
LOCUS
DEFINITION 603039012F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
mRNA sequence.
ACCESSION BI823850
VERSION BI823850.1 GI:15935400
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 954)
NTH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML1448 row: h column: 03
High quality sequence stop: 856.
```

```
FEATURES
Source
1..954
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5179826"
/lab_host="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector:
PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."
BASE COUNT 199 a 297 c 250 g 208 t
ORIGIN
```

Query Match 13.9%; Score 284.2; DB 10; Length 954;
Best Local Similarity 73.9%; Pred. No. 1.9e-24;

```
Matches 388; Conservative 0; Mismatches 133; Indels 4; Gaps 2;
QY 335 gatcgtgtgcacgtgctgctgtcctcctgtggcgtcggctcattcttctgggccaag 394
Db 198 GAGCGCATGCACGTGCTGTTCTCTTCCCTGGCGCTCTGCTCTTCTTTGTGGGCAG 257
QY 395 ctcttcagccagcaccgccacgtcttctacccatgagcccgctggtgcatgtgtggacc 454
Db 258 CTTTGTGGCAGCACCACAGATGTTTCTACCTGATGGAGCCCGCTGGCAGTGGATG 317
QY 455 accctgtgcagggcagcgcggaacgctgcacatggccgtgcgcgacatgatgcgtct 514
Db 318 ACCITCAAGCAGACACCGCCTGGATGCTGTCACATGGCTGTGCGGGATCTGATACGGGCC 377
QY 515 atctttttgtcgacatggacgtgttttgatgcctacatgccacagag---ccgaacacgtg 571
Db 378 GTCTTCTTGTGGACATGAGCGTCTTTGATGCTTACATGGAACCTGGTCCCGGAGACAG 437
QY 572 tccgcctttttcaactgggcaacgagccgcgcgtgtgctgcgcgcgcgcctgcagcgcc 631
Db 438 TCCAGCCTCTTTCAGTGGGAGAACAGCCGCGCCTGTGTCTTCTGCACCTGCTGTGACATC 497
QY 632 ttcccccaggccaccatcagcaagcaggacgtatgcaagacactgtgcacgcgccagcca 691
Db 498 ATCCACACAGATGAATCATCCCCGGGCTCACTGCAGGCTCCTGTGACATCAACAGCCC 557
QY 692 ttcagcctggcccgaggcctgcgcctcctacagccacgctggtgtcctcaaggaggtgcgc 751
Db 558 TTTGAGGTGGTGGAGAGAGCGCTGCGCTCTTACAGCCACGTTGTTCAAGGAGGTGCCG 617
QY 752 ttctcaacctgcaggtgctctaccgcgtgctcagcgaccccgctcaacctgcgcctc 811
Db 618 TTTCAACCTGCAGTCCCTCTACCCGCTGCTGAAAGACCCCTCCCTCAACCTGCATATC 677
QY 812 gtgcacc-tggtgcgcgacccgcgcgcgtgctgcgcgcgcgcgcgcgcgcgcgcgcgcgc 855
Db 678 GTGCACCTTGGTCCGGGACCCCGGCGCGTGTTCGGTTCCCGAGAGA 722
```

RESULT 12

CNS04QFN/c

LOCUS

DEFINITION

CNS04QFN

849 bp DNA

Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 129006 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION

AL302540

VERSION

AL302540.1 GI:8181872

KEYWORDS

GSS; genome survey sequence.

SOURCE

Tetraodon nigroviridis.

ORGANISM

Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE

1 (bases 1 to 849)

AUTHORS

Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

TITLE

Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 849)

AUTHORS

Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.

TITLE

Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL

Unpublished

REFERENCE

3 (bases 1 to 849)

AUTHORS

Genoscope.

TITLE

Direct Submission

JOURNAL

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT

This sequence is a single read and was generated as part of a large


```

Db 662 tttaggtggtggaagccctgccgtcctacagccacgtggtcctcaaggaggtgcgc 721
QY 752 ttcttcaacctgcagtgctctacccgctgctcaagcagaccccgcgctcaacctgcgcac 811
Db 722 ttcttcaacctgcagtcctctacccgctgctgaaagacccctccctcaacctgcatac 781
QY 812 gtgcacctggtgcgcgaccccgggcccgctgctgcgtcccgagggcggggcccgcata 871
Db 782 gtgcacctggtgcgggaccccgggcccggtgtccgttcccgcagaaacacaaaaggagat 841
QY 872 ctggcacgcgacaacggcatcgtgctgggcacccaacgcaagtgggtggaggccgacct 931
Db 842 ctcatgattgacagtcgcattgtgatgggcagcattgagcagaaactcaagaaggagac 901
QY 932 cacctgcgcctgattgcgaggtgtgccgcagccacgtgcgcacgcccaggccgcaca 991
Db 902 caacctactatgtgacaggtcatctgccaaagccagcctggagatctacaagaccac 961
QY 992 ctcaagccgcccaccttctgcgcggccgctaccgccttggtgcgcttcgagacctggcg 1051
Db 962 cagtccttgcceaagccctgcagaaacgctaccctgctgtgcgtatgaggacctggct 1021
QY 1052 cgggagccgctggcagagatccgcgcactctacgccttcaccggcctgacccctcacgcca 1111
Db 1022 cgagccctgtggcccagacttcccgaatgtatgaattcgtggattggaattcttgccc 1081
QY 1112 cagctcgaggccctggatccacaacatcacccacgggtgcggggatcggcaagccaatcgag 1171
Db 1082 catctcagacctgggtgcataacatcaccccgaggcaaggccatgggtga-----ccac 1135
QY 1172 gccttcataacttctgttaggaatgcgcgcaacgctctcccagggcctgcgcacgcgttg 1231
Db 1136 gctttcacacaaaatgccaggatgcccttaattgtctcccaggcttggcgtggtcttg 1195
QY 1232 cccctcactaagatccctgcgcgtgcaggaggtgtgcgcgcgcgcgtgcagctgctgggc 1291
Db 1196 cccatgaaaaggttctcgcacttcagaaagcctgtggcgatgccatgaatttgcctgggc 1255
QY 1292 taccggcctgtgtactctgcggaccagcagcgtgacctcaccttgatctggtg 1345
Db 1256 taccgcacgtcagatctgaacaagaacagagaaacctgttgcgtgatatctctg 1309

RESULT 2
US-08-899-514-1
; Sequence 1, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
; APPLICANT: HABUCHI, OSAMI
; APPLICANT: FUKUTA, MASAKAZU
; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
; TITLE OF INVENTION: FOR THE POLYPEPTIDE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,514
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DANIEL E ALTMAN
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714 760 0404
; TELEFAX: 714 760 9502
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2156
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Human
; TISSUE TYPE: Fetal brain
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 147..1583
; IDENTIFICATION METHOD: S
US-08-899-514-1

Query Match 11.5%; Score 235.2; DB 2; Length 2156;
Best Local Similarity 54.0%; Pred. No. 4.6e-37;
Matches 589; Conservative 0; Mismatches 468; Indels 33; Gaps 4;

QY 308 gggccctcatccccagccggcgaggatcgtgtcacgtgctggtgctgcctcgtgg 367
Db 510 GAGCCGCCAGACCGGCCGTGGCGGGCCCCGCCACGTCGTGCTCATGGCCACCAG 569
QY 368 cgtcggggtcactccttcttgggccagctcttcagccagcaccccgacgtcttctacctg 427
Db 570 CGCACCCGGCTCCTCGTTCGTGGGGGAGTTCTTCAACCAGAGGGGCAACATCTTCTACCTC 629
QY 428 atggagcccgctggcctgtgtgtggaccacctgtc-----gcaggggcagcgcg 475
Db 630 TTCGAGCCCGCTGTGGCACATCGAGCGGCACAGTGTCTCTCGAGCCCGGGGGGCCAACGCC 689
QY 476 gcaacgctgcacatggccgtgcgcgacctgatgcgctctatcttcttttgcgcacatggac 535
Db 690 GCGGGCTCGGCCCTGGTGTACCCGGACGTCGCTCAAGCAGCTCTTCCTGTGGCACCTGTAC 749
QY 536 gtgtttgatgcctaca-----tgccaacagagccgaaacctgtcgcgcttttcaac 586
Db 750 GTGCTGGAGCACTTCATCACGCCGCTGCCCGAGGACCACCTGACTCAGTTCATGTTCCGC 809
QY 587 tgggcaacgagccgcgcgtgtgctgcgcgcgcgcctgcagcgcctttcccccaggcacc 646
Db 810 CGGGGCTCCAGCCGCTCCCTGTGGAGGACCCCGCTGTACGCCCTTCGTCGAAGAAGGTC 869
QY 647 atcagcaagcagcagcgtatgcaagacactgtgcacgcggcagccattcagcctggcccgg 706
Db 870 TTCGAGAAGTACCAGTCAAGAACCCCGCGCTGGGCCCTCAACGTCAGCTGGCCGCA 929
QY 707 gaggcctgcgcctcctacagccacgtggtgtcgaaggaggtgcgcttcttcaacctgcag 766
Db 930 GAGGCCTGCCCGCAAGGAGCACATGGCCCCCTCAAGGGCGGTGGCATCCGGCAGCTGGAG 989
QY 767 gtgctctacccgctgcacgacccccgcgtcaacctgcgcacatcgtgcacctggtgcgc 826
Db 990 TTCCTGCAGCCGCTGGCCGAGGACCCCCCGCTGACCTGCGCTCATCCAGCTGGTGCGC 1049
QY 827 gaccgcgggcccgtgctgcgtcccgaggcgggggcccgatactactggcacgcgacaaac 886
Db 1050 GACCCCGGGCGCTGCTGGCCTCGCGCATGTTGGCTTCGCGGAAGTATAAGACCTGG 1109
QY 887 ggcacgtgctgggcaccaaagcaagtgggtggaggccgacctcacctggcctgatt 946
Db 1110 AAGAAGTGGCTGGACGACGAGGGCCAGGACCGCCCTGAGGGAAGAGGAGGTGCAGCGGCTG 1169

```



```
QY 1795 aaacccctctactaaaaatgcaaaaattagtcggcggtggtggcagactcctgtaat 1854
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1261 AAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGGTGGTACGCCCTGTAAT 1202
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1855 ccagctacttaggaggtgaggtgggaaaaatcacttggaactocaaaggtggagttgca 1914
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1201 CCCAGCTACTCCGAGGCTGAGGAGGAGGAGAAATCATTTGAGCCTGGGAGGTGGAGTTGCA 1142
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1915 gtaagctgaaatcatgccactgcacccctagcttggtggcaaaagcaaaactctatcaaaa 1974
|| || || || || || || || || || || || || || || || || || || || ||
Db 1141 GTGGCCGGAATAGTGCCACTACACTCCAGCCTGGGTGACACAGCAAGACTCTGTCTCAA 1082
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1975 aaataattaataa 1987
||||| | ||
Db 1081 AAATAAAAAATAAAA 1069
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 12
US-08-800-929A-5/c
; Sequence 5, Application US/08800929A
; Patent No. 6133437
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF
; TITLE OF INVENTION: IAPS AND NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,929A
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/030,590
; FILING DATE: 14-NOV-1996
; APPLICATION NUMBER: 60/017,354
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bieker-Brady, Kristina
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 07891/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdNA
US-08-800-929A-5

Query Match 9.8%; Score 200.2; DB 3; Length 6669;
Best Local Similarity 79.6%; Pred. No. 3.4e-30;
Matches 249; Conservative 0; Mismatches 63; Indels 1; Gaps 1;
```

```
Matches 249; Conservative 0; Mismatches 63; Indels 1; Gaps 1;
QY 1676 aattaaaatgtgtgccaggccgggtcggtggctcatgcctgttaatccagcattttgaga 1735
|| || || || || || || || || || || || || || || || || || || || ||
Db 1381 AAAAAACAGAGGCCTCTGCCTGGTGCAGTGGCTCAGCCCTGTAATCCACGANTTTGGGA 1322
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1736 ggctgagggggtggtgatcacctgaggtcagga-ttcaaaaccagcctggccaacatagt 1794
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1321 GGCCAAAGACAAGTGGATCACTTGAGGTGAGGAGTTCAAAACCAGCCTGGCAAAATGGTG 1262
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1795 aaacccctctactaaaaatgcaaaaattagtcggcggtggtggcagactcctgtaat 1854
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1261 AAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGGTGGTACGCCCTGTAAT 1202
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1855 ccagctacttaggaggtgaggtgggaaaaatcacttggaactocaaaggtggagttgca 1914
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1201 CCCAGCTACTCCGAGGCTGAGGAGGAGGAGAAATCATTTGAGCCTGGGAGGTGGAGTTGCA 1142
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1915 gtaagctgaaatcatgccactgcacccctagcttggtggcaaaagcaaaactctatcaaaa 1974
|| || || || || || || || || || || || || || || || || || || || ||
Db 1141 GTGGCCGGAATAGTGCCACTACACTCCAGCCTGGGTGACACAGCAAGACTCTGTCTCAA 1082
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1975 aaataattaataa 1987
||||| | ||
Db 1081 AAATAAAAAATAAAA 1069
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 13
US-09-617-053A-5/c
; Sequence 5, Application US/09617053A
; Patent No. 6300492
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
; TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
; TITLE OF INVENTION: DISEASE
; FILE REFERENCE: 07891/009003
; CURRENT APPLICATION NUMBER: US/09/617,053A
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/800,929
; PRIOR FILING DATE: 1997-02-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 6669
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variation
; LOCATION: (3677)...(3951)
; OTHER INFORMATION: n can be any nucleotide
US-09-617-053A-5

Query Match 9.8%; Score 200.2; DB 4; Length 6669;
Best Local Similarity 79.6%; Pred. No. 3.4e-30;
Matches 249; Conservative 0; Mismatches 63; Indels 1; Gaps 1;
```

```
QY 1676 aattaaaatgtgtgccaggccgggtcggtggctcatgcctgttaatccagcattttgaga 1735
|| || || || || || || || || || || || || || || || || || || || ||
Db 1381 AAAAAACAGAGGCCTCTGCCTGGTGCAGTGGCTCAGCCCTGTAATCCACGANTTTGGGA 1322
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1736 ggctgaggcggtggtgatcacctgaggtcagga-ttcaaaaccagcctggccaacatagt 1794
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1321 GGCCAAAGACAAGTGGATCACTTGAGGTGAGGAGTTCAAAACCAGCCTGGCAAAATGGTG 1262
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1795 aaacccctctctactaaaaatgcaaaaattagtcggcggtggtggcagactcctgtaat 1854
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

Db 1261 AAACCCCATCTCTACTAAAAATACAAAAAATTAGCTGGGTGGGTACGCGCCTGTAAAT 1202
QY 1855 ccaagctacttaggaggtgaggtgggaaaaatcacttggactccaaaggttgaggttgca 1914
Db 1201 CCCAGCTACTCCGGAGGCTGAGGCAGGAGAAATCATTTGAGCCTGGGAGGTGGAGTTGCA 1142
QY 1915 gtaagctgaaatcatgccactgcaccctagcttgggtggcaaaagcaaaactctatcaaaa 1974
Db 1141 GTGGCCGAAATAGTGCCACTATACACTCCAGCCTGGGTGACACAGCAAGACICTGTCTCAAA 1082
QY 1975 aaataaattaataa 1987
Db 1081 AAATAAAATAAAA 1069

RESULT 14
US-08-133-629-8
; Sequence 8, Application US/08133629
; Patent No. 5597694
; GENERAL INFORMATION:
; APPLICANT: Munroe, David J.
; APPLICANT: Housman, David E.
; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/133,629
; FILING DATE: 07-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Greer, Helen
; REGISTRATION NUMBER: 36,816
; REFERENCE/DOCKET NUMBER: M0828/7001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX: 92-1742 EZEKIEL
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-133-629-8

Query Match 9.7%; Score 198.4; DB 1; Length 282;
Best Local Similarity 82.5%; Pred. No. 3.8e-30;
Matches 231; Conservative 6; Mismatches 42; Indels 1; Gaps 1;

QY 1692 ggcggggtgcggtgcctgtaatcccagcattttgagaggtgagcggtgga 1751
Db 1 GGCTGGCGGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGA 60
QY 1752 tcacctgaggtcagga-ttcaaaaaccagcctggccaacatagtgaaacccccctctact 1810
Db 61 TCACCTGAGGTCAGGAGTCAAGACCAGCCTGGCCAAACATGGTGAAACCCCGTCTCTACT 120
QY 1811 aaaaatgcaaaaattagtcggcggtggtggccacactcctgtaatcccagctacttaggag 1870
Db 121 AAAAATACAAAATTAGCCGGGCGTGGTGGCGCGGCGCTGTAAATCCCAGCTACTCGGGAG 180

QY 1871 gctgaggtgggaaaaatcacttggactccaaaggttgaggttgcaagctgaaatcatg 1930
Db 181 GCTGAGGCAGGAGAAATCGCTTGAAACCCAGGAGGTGGAGGYTGCAGTGAGCCGWGATCGCG 240
QY 1931 ccactgcaccctagcttgggtggcaaaagcaaaactctatc 1970
Db 241 CCACTCGACTCCAGCCTGGGCAACAGAGYRAGACTCYRTC 280
RESULT 15
US-09-608-285A-42/c
; Sequence 42, Application US/09508285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608,285A
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583,231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 14747
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (13641)
; OTHER INFORMATION: n = adenosine or guanine or cytosine or thymidine
US-09-608-285A-42

Query Match 9.7%; Score 198; DB 4; Length 14747;
Best Local Similarity 82.4%; Pred. No. 1.1e-29;
Matches 239; Conservative 0; Mismatches 50; Indels 1; Gaps 1;

QY 1692 ggcggggtgcggtgcctgtaatcccagcattttgagaggtgagcggtgga 1751
Db 10298 GGCAGGCGCGGTGGTCCACGCTGTAATCCTAGCACITTTGGGAGGCTGAGGCGGTGGA 10239
QY 1752 tcacctgaggtcagga-ttcaaaaaccagcctggccaacatagtgaaacccccctctact 1810
Db 10238 TCACCTGAGGTCAGGAGTTGAGATCAGCCTGGCCCAACGTTGGTGAAACCCCATCTCTACT 10179
QY 1811 aaaaatgcaaaaattagtcggcggtggtggccacactcctgtaatcccagctacttaggag 1870
Db 10178 AAAAATACAAAATTAGCACGCGCGTGGTGGTGTACCGCTGTAATCCAGCTACTCGGGAG 10119
QY 1871 gctgaggtgggaaaaatcacttggactccaaaggttgaggttgcaagctgaaatcatg 1930

Db 10118 GCTGAGGCAGGAGATCGCTTGAACCTGGAAGGCAGAGGTTGCAGTGAGCCGAGATCTCA 10059

Qy 1931 ccactgcaccctagcttgggtggcaagcaaaactctatcaaaaaataa 1980

Db 10058 CCACTGCACCTCCAGCCCTGGGTGACAGAGCAAGACTCTGTCTCAAAAAAAA

Search completed: September 24, 2002, 09:30:24
Job time: 12113 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 19:04:23 ; Search time 1398.08 Seconds
(without alignments)
1440.505 Million cell updates/sec

Title: US-09-593-828-4
Perfect score: 1173
Sequence: 1 atgtggctgccacggttctc.....gctgggcacgcctgactga 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1173	100.0	2044	22	AAD02699
2	1173	100.0	2170	22	AAD02698
3	1173	100.0	160552	22	AAD02697
4	1027.2	87.6	2988	21	AAC76156
5	924.6	78.8	1647	24	AAD24670
6	924.6	78.8	1694	22	AAD02700
7	745.2	63.5	1937	24	AAS16948
8	745.2	63.5	1989	22	AAD02696
9	400	34.1	1979	22	AAK94229
					Human glycosyl sul
					Human glycosyl sul
					Human glycosyl sul
					Human ORFX ORF1711
					Human drug metabol
					Human glycosyl sul
					Murine intestinal-
					Mouse glycosyl sul
					Human full-length

10	400	34.1	2032	20	AAZ20792	Human glycosyl sul
11	398.4	34.0	1333	24	AAS16947	Human L-selectin s
12	377	32.1	2065	21	AAZ94211	Human transferrase
13	354	30.2	1926	20	AAZ20793	Mouse glycosyl sul
14	297.6	25.4	877	22	AAK91803	Human cDNA 5'-end
15	297.6	25.4	877	22	AAK93921	Human cDNA clone r
16	235.2	20.1	2156	19	AAV21200	Glycosaminoglycan
17	196.8	16.8	2354	18	AAT45037	Chick chondroitin
18	153.8	13.1	3029	22	AAH17922	Human cDNA sequenc
19	148.8	12.7	2393	24	AAS61699	Lung small cell ca
20	148.8	12.7	2409	20	AAK87821	Human N-acetylgluc
21	139	11.8	1458	19	AAV36418	Keratan sulphate 6
22	133	11.3	2150	20	AAK87820	Mouse N-acetylgluc
23	64.6	5.5	43280	18	AAT80413	Tylactone synthase
24	57.8	4.9	731	20	AAZ24563	Human lung tumor a
25	57.8	4.9	731	21	AAC65802	Human lung cancer-
26	57.8	4.9	1644	23	AAS54247	Pseudomonas aerugi
c 27	57.2	4.9	30001	18	AAT61016	Total DNA sequence
c 28	57.2	4.9	30001	20	AAK05110	S. aureofaciens DN
29	56.8	4.8	303	24	AAS61698	Lung small cell ca
30	56.8	4.8	303	24	AAS61848	Lung small cell ca
c 31	55	4.7	114955	20	AAK53491	Human adenosine A1
32	54.8	4.7	58857	21	AAK58471	Nucleotide sequenc
33	54.6	4.7	4689	21	AAZ87299	S. venezuelae macr
34	54.6	4.7	36778	21	AAZ87318	S. venezuelae pik
35	54.6	4.7	37948	21	AAZ87285	Nucleotide sequenc
36	54.6	4.7	38506	21	AAA75633	Recombinant cosmid
37	54.6	4.7	38506	21	AAZ56001	FLGA insert stabil
c 38	54	4.6	795	19	AAV55830	Micromonospora DNA
39	54	4.6	109519	22	AAS08693	Epstein Barr Virus
c 40	53.8	4.6	1925	20	AAK90924	Amycolatopsis medi
41	53.4	4.6	53789	19	AAV21187	Streptomyces nous
42	53.4	4.6	65140	22	AAD17184	Streptomyces nous
43	53.4	4.6	125401	22	AAD17186	Pseudomonas fluore
44	53.2	4.5	1683	22	AAD03820	Epstein Barr virus
c 45	52.8	4.5	1926	21	AAA50254	

ALIGNMENTS

RESULT 1
AAD02699
ID AAD02699 standard; cDNA; 2044 BP.

AC AAD02699;

XX 02-MAY-2001 (first entry)

DT Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.

DE Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1; ss.

XX Homo sapiens.

OS Key Location/Qualifiers
CDS 218..1390

FT /*tag= a
FT /product= "Human glycosyl sulfotransferase-4alpha
FT (GST-4alpha)"
FT /note= "CDS is specifically claimed as SEQ ID NO: 4
FT in claim 6 (page no: 41) of the specification"

XX WO200106015-A1.

Db	48076	cggtgtgcacgtgtgtgtgtgtcctcctcgttggcgtcgggtcctccttcttcttggccagctc	48133
QY	181	ttcagccagcaccccgacgtcttctacacctgatgagcccgctggccatgtgtgtggacaacc	240
Db	48136	ttcagccagcaccccgacgtcttctacacctgatgagcccgctggccatgtgtgtggacaacc	48195
QY	241	ctgtcgcagggcagcgcggcaacgcctgcacatgcccgtgcgcgacctgatgcgtctatc	300
Db	48196	ctgtcgcagggcagcgcggcaacgcctgcacatgcccgtgcgcgacctgatgcgtctatc	48255
QY	301	ttttttgtgcgacatggacgtgttttgatgcctacatgccacagagccgaacctgtccgcc	360
Db	48256	ttttttgtgcgacatggacgtgttttgatgcctacatgccacagagccgaacctgtccgcc	48315
QY	361	tttttcaactgggcaacgagccgcgcgtgtgtctgcgccgccgcctgcagcgcctttccc	420
Db	48316	tttttcaactgggcaacgagccgcgcgtgtgtctgcgccgccgcctgcagcgcctttccc	48375
QY	421	cgaggcaccatcagcaagcaggacgtatgcaagacacctgtgcacgcggcagccattcagc	480
Db	48376	cgaggcaccatcagcaagcaggacgtatgcaagacacctgtgcacgcggcagccattcagc	48435
QY	481	ctggcccgggaggcctgccgtcctacagccacgttgtgtcctcaaggaggtgcgtctctc	540
Db	48436	ctggcccgggaggcctgccgtcctacagccacgttgtgtcctcaaggaggtgcgtctctc	48495
QY	541	aacctgcaggtgtctctaccgcctgtcctcagcgaccccgccgcctcaacctgcgcac	600
Db	48496	aacctgcaggtgtctctaccgcctgtcctcagcgaccccgccgcctcaacctgcgcac	48555
QY	601	ctggtgcgcgaccccgggccgtgtgcgtcctcccgaggcgggcgcccgatactggca	660
Db	48556	ctggtgcgcgaccccgggccgtgtgcgtcctcccgaggcgggcgcccgatactggca	48615
QY	661	cgcgacaacggcatcgtgtgtgggcaacaaaggcaagtgggtgaggccgacccctcacctg	720
Db	48616	cgcgacaacggcatcgtgtgtgggcaacaaaggcaagtgggtgaggccgacccctcacctg	48675
QY	721	cgcctgattcgcgaggtgtgcgcgagccacgtgcgcacatgcgcgagggccgacactcaag	780
Db	48676	cgcctgattcgcgaggtgtgcgcgagccacgtgcgcacatgcgcgagggccgacactcaag	48735
QY	781	ccgccacccttccctgcgcggccgctacccgctgtgcgttcgaggaacctggcgcgggag	840
Db	48736	ccgccacccttccctgcgcggccgctacccgctgtgcgttcgaggaacctggcgcgggag	48795
QY	841	ccgctggcagagatccgcgcactctacgccttcacccggtcaccctcagccacagctc	900
Db	48796	ccgctggcagagatccgcgcactctacgccttcacccggtcaccctcagccacagctc	48855
QY	901	gaggccctggatccacaacatcacccacgggtcggggatcggcaagccaatcgaggccttc	960
Db	48856	gaggccctggatccacaacatcacccacgggtcggggatcggcaagccaatcgaggccttc	48915
QY	961	catacttcgtctaggaatgcgcgcaacgtctcccaggcctggcgccacgcgttgcccttc	1020
Db	48916	catacttcgtctaggaatgcgcgcaacgtctcccaggcctggcgccacgcgttgcccttc	48975
QY	1021	actaagatccctgcgcgtgcaggaggtgtgcgcggcgcgctgcagctgctgggctaccgg	1080
Db	48976	actaagatccctgcgcgtgcaggaggtgtgcgcggcgcgctgcagctgctgggctaccgg	49035
QY	1081	cctgtgtactctgcggaccagcagcgtgacctcacccctggatctggtgtgcaacgaggc	1140
Db	49036	cctgtgtactctgcggaccagcagcgtgacctcacccctggatctggtgtgcaacgaggc	49095
QY	1141	ccagaccacttcagctggggcatcgcctgactga	1173
Db	49096	ccagaccacttcagctggggcatcgcctgactga	49128

AAC76156 standard; cDNA; 2988 BP.

AAC76156;

08-FEB-2001 (first entry)

Human ORFX ORF1711 polynucleotide sequence SEQ ID NO:3421.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnery; antipsoptic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M;

WPI; 2000-602362/57.
P-PSDB; AAB41947.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease-

Claim 5; Page 2597-2599; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnery; antispasmodic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressive; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 19:02:45 ; Search time 1398.08 Seconds
(without alignments)
2510.138 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggctcgaggtccactgtgct.....aaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2044	100.0	2044	22	AAD02699
2	1973	96.5	2170	22	AAD02698
3	1827.2	89.4	2988	21	AAC76156
4	1753.6	85.8	160552	22	AAD02697
5	1111.4	54.4	1694	22	AAD02700
6	947	46.3	1647	24	AAD24670
7	751.2	36.8	1989	22	AAD02696
8	750	36.7	1937	24	AAS16948
9	400	19.6	1979	22	AAK94229
					Human glycosyl sul
					Human glycosyl sul
					Human ORFX ORF1711
					Human glycosyl sul
					Human glycosyl sul
					Human drug metabol
					Mouse glycosyl sul
					Murine intestinal-
					Human full-length

10	400	19.6	2032	20	AAZ20792	Human glycosyl sul
11	398.4	19.5	1333	24	AAZ16947	Human L-selectin s
12	377	18.4	2065	21	AAZ94211	Human transfere
13	354	17.3	1926	20	AAZ20793	Mouse glycosyl sul
14	297.6	14.6	877	22	AAK91803	Human cDNA 5'-end
15	297.6	14.6	877	22	AAK93921	Human cDNA clone r
16	235.2	11.5	2156	19	AAV21200	Glycosaminoglycan
17	222	10.9	12503	22	AAZ31470	Human DNA for a no
18	216.4	10.6	107820	22	AAZ16230	Human ATP-binding
19	214.8	10.5	32249	22	ABA17155	Human nervous syst
20	211.8	10.4	24843	24	AAZ17764	Human Genomic DNA
21	211	10.3	1043	21	AAZ14459	Human interleukin-
22	211	10.3	1601	21	AAZ21313	Human low adenosin
23	211	10.3	1601	21	AAZ35191	Human adenosine re
24	211	10.3	1601	21	AAZ46814	Interleukin-10 (IL
25	211	10.3	1618	12	AAQ10207	pH15C insert conta
26	211	10.3	1618	14	AAQ46958	Human cytokine syn
27	211	10.3	1618	21	AAZ86905	Human CSIF coding
28	211	10.3	1618	22	AAF90433	Human cytokine syn
29	211	10.3	1618	22	AAF82588	Human CSIF cDNA.
30	211	10.3	1618	22	AAF83186	Human CSIF protein
31	211	10.3	1645	21	AAZ58659	Human interleukin-
32	211	10.3	2797	22	AAH18187	Human cDNA sequenc
33	211	10.3	15630	21	AAF21317	Human low adenosin
34	211	10.3	15630	21	AAA35195	Human adenosine re
35	209.8	10.3	32152	22	ABA08132	Human ovarian and
36	209.2	10.2	23989	22	AAK72555	Human immune/haema
37	209	10.2	14209	22	AAK89137	Human digestive sy
38	208.8	10.2	21477	22	AAK66626	Human immune/haema
39	208.8	10.2	21480	22	AAK66625	Human immune/haema
40	208.8	10.2	27571	22	ABA19222	Human nervous syst
41	208.8	10.2	27572	22	ABA19221	Human nervous syst
42	208.4	10.2	15589	22	AAK80921	Human immune/haema
43	208.4	10.2	15610	22	AAZ07072	Human reproductive
44	208.4	10.2	15610	22	AAK80920	Human immune/haema
45	208.2	10.2	4125	22	AAK84260	Human immune/haema

ALIGNMENTS

RESULT 1
AAD02699
ID AAD02699 standard; cDNA; 2044 BP.

XX AAD02699;

AC AAD02699;

XX 02-MAY-2001 (first entry)

DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.
XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenailitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1; ss.

OS Homo sapiens.

XX
FH Key Location/Qualifiers
FT CDS 218..1390
FT /*tag= a
FT /product= "Human glycosyl sulfotransferase-4alpha
FT (GST-4alpha)"
FT /note= "CDS is specifically claimed as SEQ ID NO: 4
FT in claim 6 (page no: 41) of the specification"
XX

PN WO200106015-A1.

XX 25-JAN-2001.
PD
XX
PF 19-JUL-2000; 2000WO-US19741.
XX
PR 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Rosen SD, Lee JK, Hemmerich S;
XX
DR WPI; 2001-138471/14.
DR P-PSDB; AAY72639.
XX
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications -
XX
PS Claim 6; Fig 1; 128pp; English.
XX
CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
CC alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand. GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
CC Note: The present sequence is also shown in sequence listing (page
CC no: 56) but lacks four nucleotides at its 3' end.
XX
SQ Sequence 2044 BP; 422 A; 645 C; 570 G; 407 T; 0 other;

Query Match 100.0%; Score 2044; DB 22; Length 2044;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggctcgaggtccactgtgctgaatgtaagttaagtctccttatcagaaagctcccagtgaggaac 60
|||
Db 1 ggctcgaggtccactgtgctgaatgtaagttaagtctccttatcagaaagctcccagtgaggaac 60
|||
QY 61 tggctctctggagactctgtgtggcatagagtgtattcaccacaccttaagaagacctctgg 120
|||
Db 61 tggctctctggagactctgtgtggcatagagtgtattcaccacaccttaagaagacctctgg 120
|||
QY 121 ctttcctggaacacagatgtcgagacatctcccatggattgtgatcagcgtttgcagctc 180
|||
Db 121 ctttcctggaacacagatgtcgagacatctcccatggattgtgatcagcgtttgcagctc 180
|||
QY 181 tcccagcagccctggacggtggccccccagcccgccatgtggctgccacggttctccag 240
|||
Db 181 tcccagcagccctggacggtggccccccagcccgccatgtggctgccacggttctccag 240
|||
QY 241 caagacagtgacagtgtctctcctgtggcacagaccactgcctcctgtctcttcacatctc 300
|||
Db 241 caagacagtgacagtgtctctcctgtggcacagaccactgcctcctgtctcttcacatctc 300
|||
QY 301 ccggccagggccctcatccccagccggcgaggatcgtgtgcacgtgctggtgctgc 360
|||
Db 301 ccggccagggccctcatccccagccggcgaggatcgtgtgcacgtgctggtgctgc 360
|||
QY 361 ctcgtagcgctcgggctcactccttcttggccagctcttcagccagcaccgccagcttct 420
|||

Db 361 ctcgtagcgctcgggctcactccttcttgggccaagtcttctcagccagcaccgccagcttctt 420
QY 421 ctacctgatggagcccgctggcatgtgtggaccacctctgcagggcagcggcgaac 480
|||
Db 421 ctacctgatggagcccgctggcatgtgtggaccacctctgcagggcagcggcgaac 480
|||
QY 481 gctgcacatggcctgtgcgcacctgatgcgctctatcttcttctgtgcacatggaactgtt 540
|||
Db 481 gctgcacatggcctgtgcgcacctgatgcgctctatcttcttctgtgcacatggaactgtt 540
|||
QY 541 tgatgcctacatgccacagagccgaacctgtccgcttttcaactgggccaacgagccg 600
|||
Db 541 tgatgcctacatgccacagagccgaacctgtccgcttttcaactgggccaacgagccg 600
|||
QY 601 cgcgctgtctgcgcgcctgcagcgccttcccccagggcaccatcagcaagcagga 660
|||
Db 601 cgcgctgtctgcgcgcctgcagcgccttcccccagggcaccatcagcaagcagga 660
|||
QY 661 cgtatgcaagacactgtgcaagcggcagccattcagcctggccgggagcctgcgcctc 720
|||
Db 661 cgtatgcaagacactgtgcaagcggcagccattcagcctggccgggagcctgcgcctc 720
|||
QY 721 ctacagccacgtgtgtctcaaggaggtgcgttcttcaacctgcaggtgtctctacccgt 780
|||
Db 721 ctacagccacgtgtgtctcaaggaggtgcgttcttcaacctgcaggtgtctctacccgt 780
|||
QY 781 gctcagcagccccgcgtcaacctggcgcacctgtgcacctgtgtgcgcgcgggcgt 840
|||
Db 781 gctcagcagccccgcgtcaacctggcgcacctgtgcacctgtgtgcgcgcgggcgt 840
|||
QY 841 gctcgcctcccgaggcggcggccgatactggcacgcgacacagcgcacctgtgtggtg 900
|||
Db 841 gctcgcctcccgaggcggcggccgatactggcacgcgacacagcgcacctgtgtggtg 900
|||
QY 901 caccacggcaagtgggtggaggccgacctcacctgcgcctgattcggaggtgtgctg 960
|||
Db 901 caccacggcaagtgggtggaggccgacctcacctgcgcctgattcggaggtgtgctg 960
|||
QY 961 cagccacgtgcgcacctgcggaggccgcacactcaagccaccttctcctgcgcggcgt 1020
|||
Db 961 cagccacgtgcgcacctgcggaggccgcacactcaagccaccttctcctgcgcggcgt 1020
|||
QY 1021 ctaccgcctgtgtgcgtcttcgaggacctggcggggagccgctggcagagatccgcgcact 1080
|||
Db 1021 ctaccgcctgtgtgcgtcttcgaggacctggcggggagccgctggcagagatccgcgcact 1080
|||
QY 1081 ctacgccttcacccgctgacctcacgccaagcgcacagctcagcctggatccacaacatcac 1140
|||
Db 1081 ctacgccttcacccgctgacctcacgccaagcgcacagctcagcctggatccacaacatcac 1140
|||
QY 1141 ccacgggtcggggtatcggcaagccaatcgaggccttcacactccttcctaggaatgcgcg 1200
|||
Db 1141 ccacgggtcggggtatcggcaagccaatcgaggccttcacactccttcctaggaatgcgcg 1200
|||
QY 1201 caacgtctcccaggcctggcgccacgcgttgccttcaagatacctgcgcgtgcaagga 1260
|||
Db 1201 caacgtctcccaggcctggcgccacgcgttgccttcaagatacctgcgcgtgcaagga 1260
|||
QY 1261 ggtgtgcgcggcgcgtgcagctgtgggtaccgctggtgtactctggtgaccagca 1320
|||
Db 1261 ggtgtgcgcggcgcgtgcagctgtgggtaccgctggtgtactctggtgaccagca 1320
|||
QY 1321 gctgacctcacccctggatctgtgtgtgctgccagagggccacacttcagctgggcactc 1380
|||
Db 1321 gctgacctcacccctggatctgtgtgtgctgccagagggccacacttcagctgggcactc 1380
|||
QY 1381 gctgacctgagaactctgggcttagagcaagccccgaactgtggtgcgagggccagga 1440
|||
Db 1381 gctgacctgagaactctgggcttagagcaagccccgaactgtggtgcgagggccagga 1440
|||
QY 1441 agcgtgcatggtggaaaaggagctggggcgcatggggaacaggtccctactatcaacc 1500
|||
Db 1441 agcgtgcatggtggaaaaggagctggggcgcatggggaacaggtccctactatcaacc 1500
|||

QY 1501 gggagtttgggtccctccctgaagtaagcaaggactgcacgtttctttctctctgatt 1560
Db 1501 gggagtttgggtccctccctgaagtaagcaaggactgcacgtttctttctctctgatt 1560
QY 1561 ctcggttttctctttagtcttctggagctgccttctcatcaggtgcactcttcatgaaa 1620
Db 1561 ctcggttttctctttagtcttctggagctgccttctcatcaggtgcactcttcatgaaa 1620
QY 1621 agcaactcttgccctacctcttctggcgaggaagtaagtaagtaagtaagtaagta 1680
Db 1621 agcaactcttgccctacctcttctggcgaggaagtaagtaagtaagtaagtaagta 1680
QY 1681 aatgtgtccaggccgggtgcggtgggtcctcatgcctgtaatcccagcattttgagagctg 1740
Db 1681 aatgtgtccaggccgggtgcggtgggtcctcatgcctgtaatcccagcattttgagagctg 1740
QY 1741 agggggttgatcacctgaggtcaggattcaaaaaccagccctggccaacatagtgaaacc 1800
Db 1741 agggggttgatcacctgaggtcaggattcaaaaaccagccctggccaacatagtgaaacc 1800
QY 1801 cctctactataaaatgcaaaaattagtcggcgctggtggcagactcctgttaatcccagc 1860
Db 1801 cctctactataaaatgcaaaaattagtcggcgctggtggcagactcctgttaatcccagc 1860
QY 1861 tacttaggaggtgaggtgggaaaatcacttggaactccaaagggtggaggttgagtaagc 1920
Db 1861 tacttaggaggtgaggtgggaaaatcacttggaactccaaagggtggaggttgagtaagc 1920
QY 1921 tgaatatcatgccactgcacacttagcttgggtggcagcaaaactctatcaaaaaataa 1980
Db 1921 tgaatatcatgccactgcacacttagcttgggtggcagcaaaactctatcaaaaaataa 1980
QY 1981 ttaataaattgttcaaaagtccctgcccgaataaaataaaataaaataaaataaa 2040
Db 1981 ttaataaattgttcaaaagtccctgcccgaataaaataaaataaaataaaataaa 2040
QY 2041 aaaa 2044
Db 2041 aaaa 2044

RESULT 2

AAD02698

ID AAD02698 standard; cDNA; 2170 BP.

XX

AC AAD02698;

XX

DT 02-MAY-2001 (first entry)

XX

DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #1.

XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection; chromosome 16q23.1; ss.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT 5'UTR 9..343

FT /*tag= a

FT CDS 344..1516

FT /*tag= b

FT /product= "Human glycosyl sulfotransferase-4alpha (GST-4alpha)"

FT /note= "CDS is specifically claimed as SEQ ID NO: 4

FT 3'UTR in claim 6 (page no: 41) of the specification"
FT 1517..2134
FT /*tag= c

XX WO200106015-A1.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US19741.

XX 20-JUL-1999; 99US-0144694.

XX 13-JUL-2000; 2000US-0593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX P-PSDB; AAY72639.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications -
Claim 6; Page 62; 128pp; English.
The present sequence is human glycosyl sulfotransferase-4alpha (GST-4 alpha) cDNA. GST-4 gene is found on chromosome 16q23.1. GST is a type 2 membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non-sulphated selectin ligand, GST and a small molecular agent that inhibits the sulphation activity of GST. GST is also useful in inhibiting a selectin mediated binding event. GST is useful in gene therapy to treat disorders such as acute or chronic inflammation, systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis, myocarditis, regional enteritis, adult respiratory distress syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection during transplantation.

XX Sequence 2170 BP; 449 A; 682 C; 612 G; 427 T; 0 other;

Query Match

Best Local Similarity 96.5%; Score 1973; DB 22; Length 2170;

Matches 1998; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY 41 gaaagctcccagtgaggaaactggtcttctggagactctgtgtgcatagagtgattcaac 100
Db 167 gaaagctcccagtgaggaaactggtcttctggagactctgtgtgcatagagtgattcaac 226
QY 101 cacttaagaagacctctggtcttctggaacacacagatgtcgagacatctcccatgatt 160
Db 227 cacttaagaagacctctggtcttctggaacacacagatgtcgagacatctcccatgatt 286
QY 161 tgtgatcagcgttgacgctctcccagccctggagcgtggccccagcccgccatg 220
Db 287 tgtgatcagcgttgacgctctcccagccctggagcgtggccccagcccgccatg 346
QY 221 tggctgccacggttctccagcaagacagtgacagtgctctctctggcacagaccctgc 280
Db 347 tggctgccacggttctccagcaagacagtgacagtgctctctctggcacagaccctgc 406
QY 281 ctctgctcttcatcatctcccggccagggccctcatcccagccggcgaggatcgt 340
Db 407 ctctgctcttcatcatctcccggccagggccctcatcccagccggcgaggatcgt 466
QY 341 gtgcacgtgctggtgctctctctggtgcgctcgggctcattcttctggccagctcttc 400
Db 467 gtgcacgtgctggtgctctctctggtgcgctcgggctcattcttctggccagctcttc 526

QY	401	agccagcaccgcagcgtctttctacctgatggagcccgcgctggcatgtgtggaccaccctg	460
Db	527	agccagcaccgcagcgtctttctacctgatggagcccgcgctggcatgtgtggaccaccctg	586
QY	461	tcgcagggcagcgcgcaacgctgcacatggcggtgcgcgacacctgatgcgtcttatcttt	520
Db	587	tcgcagggcagcgcgcaacgctgcacatggcggtgcgcgacacctgatgcgtcttatcttt	646
QY	521	ttgtgcgacatggaacgtgtttgatgcctacatgccacagagccgaaacctgtccgccttt	580
Db	647	ttgtgcgacatggaacgtgtttgatgcctacatgccacagagccgaaacctgtccgccttt	706
QY	581	ttcaactgggcaacagcgcgcgctgtgtcgcgcgcgcctcgagcgccctttccocga	640
Db	707	ttcaactgggcaacagcgcgcgctgtgtcgcgcgcgcctcgagcgccctttccocga	766
QY	641	ggcaccatcagcaagcagcgtatgcaagacactgtgcacgcggcagccattcagcctg	700
Db	767	ggcaccatcagcaagcagcgtatgcaagacactgtgcacgcggcagccattcagcctg	826
QY	701	gcccgggagccctgccctcctacagccacgtgtgtctcaaggaggtgcgctttctcaac	760
Db	827	gcccgggagccctgccctcctacagccacgtgtgtctcaaggaggtgcgctttctcaac	886
QY	761	ctgcaggtgctctacccgctgctcagcagcccccgcgtcaacctgcgcacgtgcacctg	820
Db	887	ctgcaggtgctctacccgctgctcagcagcccccgcgtcaacctgcgcacgtgcacctg	946
QY	821	gtgcgcgaccccgcgccgtgtcgcgtcccgaggcgggcgcccgatcactggcacgc	880
Db	947	gtgcgcgaccccgcgccgtgtcgcgtcccgaggcgggcgcccgatcactggcacgc	1006
QY	881	gacaacggcatcgtgctgggcacccaacggcaagtgggtggagcgcgacccctcacctg	940
Db	1007	gacaacggcatcgtgctgggcacccaacggcaagtgggtggagcgcgacccctcacctg	1066
QY	941	ctgattcgcgaggtgtgcgcgagccacacgtgcgcacgtgcgcagcgccacactcaagc	1000
Db	1067	ctgattcgcgaggtgtgcgcgagccacacgtgcgcacgtgcgcagcgccacactcaagc	1126
QY	1001	ccaccttctcgcgcgcgcgtacccgcctggtgcgttcgagggacctggcgcgagcgc	1060
Db	1127	ccaccttctcgcgcgcgcgtacccgcctggtgcgttcgagggacctggcgcgagcgc	1186
QY	1061	ctggcagagatccgcgcaactctacgccttaacgcgcctgacccctcagccacagctcg	1120
Db	1187	ctggcagagatccgcgcaactctacgccttaacgcgcctgacccctcagccacagctcg	1246
QY	1121	gcctggatccacaacatcacccacgggtcggggatcggaagccaatcgaggcccttccat	1180
Db	1247	gcctggatccacaacatcacccacgggtcggggatcggaagccaatcgaggcccttccat	1306
QY	1181	acttcgtctaggaatgcgcgcaacgtctcccaggcctggcgccacgcgttgcccttca	1240
Db	1307	acttcgtctaggaatgcgcgcaacgtctcccaggcctggcgccacgcgttgcccttca	1366
QY	1241	aagatcctgcgcgtgcaggaggtgtgcgcgcgcgtgcagctgctgggtaccggcct	1300
Db	1367	aagatcctgcgcgtgcaggaggtgtgcgcgcgcgtgcagctgctgggtaccggcct	1426
QY	1301	gtgtactctgcggaccagcagcgtgacctacccctggatctggtgctgccacgagccca	1360
Db	1427	gtgtactctgcggaccagcagcgtgacctacccctggatctggtgctgccacgagccca	1486
QY	1361	gaccacttcagctgggcacatcgccctgactgagaactctgggccttagagcaagcccgaa	1420
Db	1487	gaccacttcagctgggcacatcgccctgactgagaactctgggccttagagcaagcccgaa	1546
QY	1421	tgtggtgccagggcccaggaagcagcactgcatggtggaaaaggagctggggcgcatgggga	1480
Db	1547	tgtggtgccagggcccaggaagcagcactgcatggtggagaggagctggggcgcatgggga	1606
QY	1481	a-caggtccctactatcaaccgggagtttggggtcctccctgaagtgaagcaaggactgc	1539
Db	1607	agcaggtccctactatcaaccgggagtttggggtcctccctgaagtgaagcaaggactgc	1666
QY	1540	acgtttctttctctcctgattctcgggttttccctttgagttcttctggagctgccttctcat	1599
Db	1667	acgtttctttctctcctgattctcgggttttccctttgagttcttctggagctgccttctcat	1726
QY	1600	caggtgcactcttccatggaaaaagcaactcttggccctaccctcttctggcgaggagta	1659
Db	1727	caggtgcactcttccatggaaaaagcaactcttggccctaccctcttctggcgaggagta	1785
QY	1660	agttactgctaaattaaattaaatgtgtgccaggccggtgcggtggtcctcatgcctgtaa	1719
Db	1786	agttactgctaaattaaattaaatgtgtgccaggccggtgcggtggtcctcatgcctgtaa	1845
QY	1720	tcccagcattttgagaggctgagggcggtggatcacctgaggtcagggtcaggaaccagc	1779
Db	1846	tcccagcattttgagaggctgagggcggtggatcacctgaggtcagggtcaggaaccagc	1905
QY	1780	ctggccaacatagtgaaacccctctctactataaaaaatgcaaaaattagtcggcgctggtg	1839
Db	1906	ctggccaacatagtgaaacccctctctactataaaaaatgcaaaaattagtcggcgctggtg	1965
QY	1840	gcacactcctgtaatcccagctacttaggaggtgaggtgggaaaaatcacttggactcca	1899
Db	1966	gcacactcctgtaatcccagctacttaggaggtgaggtgggaaaaatcacttggactcca	2025
QY	1900	aaggtggaggttgacagtaagctgaaatcatgccactgccactgacccctagctgggtggcaagc	1959
Db	2026	aaggtggaggttgacagtaagctgaaatcatgccactgccactgacccctagctgggtggcaagc	2085
QY	1960	aaaactctatcaaaaaataattaataatttgttcaaaagtccgcgaaaaa	2019
Db	2086	aaaactctatcaaaaaataattaataatttgttcaaaagtccgcgaaaaa	2145
QY	2020	aaaaaaaaaaaaaaaaaaaaa	2044
Db	2146	aaaaaaaaaaaaaaaaaaaaa	2170
RESULT 3			
AAC76156			
ID	AAC76156 standard; cDNA; 2988 BP.		
XX	AAC76156;		
AC	AAC76156;		
XX	08-FEB-2001 (first entry)		
DT	Human ORFX ORF1711 polynucleotide sequence SEQ ID NO:3421.		
XX	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.		
XX	Homo sapiens.		
OS	WO200058473-A2.		
XX	05-OCT-2000.		
PN	31-MAR-2000; 2000WO-US08621.		
XX			
PD			
XX			
PF			

Db	2422	ggaagcaggtccctactatcaacccgggagtttgggggtccctccctgaagtaggcaaggac	2481
QY	1537	tgcacgtttcttctcctcgtattctcgtgtttcttcttcttgagttcttctgagcgtccttct	1596
Db	2482	tgcacgtttcttctcctcgtattctcgtgtttcttcttcttgagttcttctgagcgtccttct	2541
QY	1597	catcaggtgcactcttcattggaagcaactcttgcacctacctcttctggcgcaaggga	1656
Db	2542	catcaggtgcactcttcattggaagcaactcttgcacctacctcttctggcgcaaggga	2600
QY	1657	gtaagttactgctaaattaaattaaatgtgtgccagcgcggtgcggtggtcctcgtcctg	1716
Db	2601	gtaagttactgctaaattaaattaaatgtgtgccagcgcggtgcggtggtcctcgtcctg	2660
QY	1717	taatcccgacattttgagaggctgagcgcggtggatcacctgaggtcaggattcaaaacc	1776
Db	2661	taatcccgacattttgagaggctgagcgcggtggatcacctgaggtcaggattcaaaacc	2720
QY	1777	agcctggccaacatagtgaaacccctctctactactaaaaatgcaaaaattagtcggcgctg	1836
Db	2721	agcctggccaacatagtgaaacccctctctactactaaaaatgcaaaaattagtcggcgctg	2780
QY	1837	gtggcacactcctgtaatccagctacttaggaggctgaggtgggaataatcacttggact	1896
Db	2781	gtggcacactcctgtaatccagctacttaggaggctgaggtgggaataatcacttggact	2840
QY	1897	ccaaaggtggaggttcagtgaaatgaaatcatgccactgacccctagcttgggtggcaa	1956
Db	2841	ccaaaggtggaggttcagtgaaatgaaatcatgccactgacccctagcttgggtggcaa	2900
QY	1957	agcaaaactctatcaaaaaataataataaatttgttcaaaagtcctgccgaaaaaaa	2016
Db	2901	agcaaaactctatcaaaaaataataataaatttgttcaaaagtcctgccgaaaaaaa	2960
QY	2017	aaaaaaaaaaaaaaaaaaaaaaaaa	2044
Db	2961	aaaaaaaaaaaaaaaaaaaaaaaaa	2988
RESULT 4			
AAD02697			
ID	AAD02697 standard; DNA; 160552 BP.		
XX			
AC	AAD02697;		
XX			
DT	02-MAY-2001 (first entry)		
XX			
DE	Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.		
XX			
KW	Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;		
KW	therapy; selectin binding inhibitor; gene therapy; inflammation;		
KW	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;		
KW	polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;		
KW	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;		
KW	Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;		
KW	demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;		
KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;		
KW	asthma; hypersensitivity; rheumatic fever; tissue rejection;		
KW	chromosome 16q23.1; ds.		
XX			
OS	Homo sapiens.		
XX			
FH	Key		
FT	Location/Qualifiers		
FT	exon		
FT	32847..32922		
FT	/*tag= a		
FT	/number= 1		
FT	/label= 4a_5U4		
FT	32923..35592		
FT	/*tag= b		
FT	/cons_splice= (5'site:NO, 3'site:YES)		
FT	35593..35674		
FT	/*tag= c		
FT	/number= 2		
FT			

FT	intron	/label= 4a_5U3
FT		35675..45093
FT		/*tag= d
FT	exon	45094..45185
FT		/*tag= e
FT		/number= 3
FT		/label= 4a_5U2
FT	intron	45186..46633
FT		/*tag= f
FT		/cons_splice= (5'site:NO, 3'site:NO)
FT	exon	46634..46700
FT		/*tag= g
FT		/number= 4
FT		/label= 4a_5U1
FT	intron	46701..47938
FT		/*tag= h
FT		/cons_splice= (5'site:YES, 3'site:NO)
FT	exon	47939..49746
FT		/*tag= i
FT		/number= 5
FT		/note= "Includes 17 base pairs of 5'UTR, the ORF
FT		and all of 3'UTR"
FT	5'UTR	47939..47955
FT		/*tag= j
FT		/note= "portion of 5' untranslated region (5'UTR)"
FT	CDS	47956..49128
FT		/*tag= k
FT		/product= "Human glycosyl transferase-4alpha
FT		(GST-4alpha)"
FT	3'UTR	49129..49746
FT		/*tag= l
FT	exon	83257..83347
FT		/*tag= m
FT		/label= 4a_5U2
FT	intron	83348..96412
FT		/*tag= n
FT		/cons_splice= (5'site:NO, 3'site:NO)
FT	exon	96413..96484
FT		/*tag= o
FT		/label= 4a_5U1
FT	intron	96485..98456
FT		/*tag= p
FT		/cons_splice= (5'site:NO, 3'site:NO)
FT	exon	98457..99968
FT		/*tag= q
FT		/note= "Includes 17 base pairs of 5'UTR, the ORF
FT		and all of 3'UTR"
FT	5'UTR	98457..98473
FT		/*tag= r
FT		/note= "Portion of 5' untranslated region (5'UTR)"
FT	CDS	98474..99661
FT		/*tag= s
FT		/product= "Human glycosyl transferase-4beta
FT		(GST-4beta)"
FT	3'UTR	99662..99968
FT		/*tag= t
XX		
PN		WO200106015-A1.
XX		
PD		25-JAN-2001.
XX		
PF		19-JUL-2000; 2000WO-US19741.
XX		
PR		20-JUL-1999; 99US-0144694.
PR		13-JUL-2000; 2000US-0593828.
XX		
PA		(REGC) UNIV CALIFORNIA.
XX		
PI		Rosen SD, Lee JK, Hemmerich S;
XX		
DR		WPI; 2001-138471/14.
DR		P-PSDB; AAY72639, AAY72640.
XX		

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications -
XX Example 1; Page 62-104; 128pp; English.
XX
CC The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
CC DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
CC chromosome 16q23.1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.

XX
SQ Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other;

Query Match 85.8%; Score 1753.6; DB 22; Length 160552;
Best Local Similarity 99.1%; Pred. No. 4.2e-280;
Matches 1795; Conservative 0; Mismatches 14; Indels 3; Gaps 3;

QY 197 cgggtgccccagccgcgcagatggtgctgcaacggttctccagcaagacagtgacagtg 256
|| |||||
Db 47935 cgaaggccccagccgcgcagtggtgctgcaacggttctccagcaagacagtgacagtg 47994

QY 257 ctctctctggcacagaccacctgcctctctgtcttctcatcatctcccgccaggccctca 316
|| |||||
Db 47995 ctctctctggcacagaccacctgcctctctgtcttctcatcatctcccgccaggccctca 48054

QY 317 tccccagccggcgagagatcgtgtgcacgtgctggtgctgtctctctgctggcctcgggc 376
|| |||||
Db 48055 tccccagccggcgagagatcgtgtgcacgtgctggtgctgtctctctgctggcctcgggc 48114

QY 377 tcctctctctgggccagctcttcagccagcaccctccagcgtcttctacatgatggagccc 436
|| |||||
Db 48115 tcctctctctgggccagctcttcagccagcaccctccagcgtcttctacatgatggagccc 48174

QY 437 gcgtggcatgtgtggaccaccctgtcgcaggggcagcggcaacgctgcacatggccgtg 496
|| |||||
Db 48175 gcgtggcatgtgtggaccaccctgtcgcaggggcagcggcaacgctgcacatggccgtg 48234

QY 497 cgcgacctgatgcgtctatatcttttctgcgacatggacgtgtttgatgcctacatgccca 556
|| |||||
Db 48235 cgcgacctgatgcgtctatatcttttctgcgacatggacgtgtttgatgcctacatgccca 48294

QY 557 cagagccgaaacctgtccgccttttcaactgggcaacgagccgcgcgtgtgctcgcgcg 616
|| |||||
Db 48295 cagagccgaaacctgtccgccttttcaactgggcaacgagccgcgcgtgtgctcgcgcg 48354

QY 617 cccgcctgcagcgcctttccccagggcaccatcagcaagcaggagcatatgcaagacactg 676
|| |||||
Db 48355 cccgcctgcagcgcctttccccagggcaccatcagcaagcaggacgtatgcaagacactg 48414

QY 677 tgcacgcggcagccattcagccctgcccgggagggcctgcctgcctcctacagccacgtggtg 736
|| |||||
Db 48415 tgcacgcggcagccattcagccctgcccgggagggcctgcctgcctcctacagccacgtggtg 48474

QY 737 ctcaaggaggtgcgtcttctcaacctgcaggtgctctacccgctgctcagcgaccccgcg 796
|| |||||
Db 48475 ctcaaggaggtgcgtcttctcaacctgcaggtgctctacccgctgctcagcgaccccgcg 48534

QY 797 ctcaacctgcgcacatcgtgcacctggtgcgcgaccccgcgcgctgctgcgtcccgggag 856
|| |||||

Db 48535 ctcaacctgcgcacatcgtgcacctggtgtgcgaccccgggccgtgctgcgtccccgggag 48594
QY 857 gcggcgggccccgatactactggcacgggacacaacggcatcgtgctgggcaccaacggcaagtgg 916
|| |||||
Db 48595 gcggcgggccccgatactactggcacgggacacaacggcatcgtgctgggcaccaacggcaagtgg 48654
QY 917 gtgagggccgacccctcacctgcgcgtgattcgcgaggtgtgctgcgcgacgccacgtgcgcac 976
|| |||||
Db 48655 gtgagggccgacccctcacctgcgcgtgattcgcgaggtgtgctgcgcgacgccacgtgcgcac 48714
QY 977 gcgcgagggccgcccacactcaagcgcgcaccccttctctgcgcggcgcgtaccgcctggtgcgc 1036
|| |||||
Db 48715 gcgcgagggccgcccacactcaagcgcgcaccccttctctgcgcggcgcgtaccgcctggtgcgc 48774
QY 1037 ttgcgaggacctggcgcggggagccgctggtgcagagatccgcgcactctacgccttcacccgc 1096
|| |||||
Db 48775 ttgcgaggacctggcgcggggagccgctggtgcagagatccgcgcactctacgccttcacccgc 48834
QY 1097 ctgacctctcacgcccacagctcgaggcctggatcccaacaacatcacccacgggtcggggac 1156
|| |||||
Db 48835 ctgacctctcacgcccacagctcgaggcctggatcccaacaacatcacccacgggtcggggac 48894
QY 1157 ggcaagccaatcgaggccttccataacttctgttaggaatgcgcgaacgtctccccaggcc 1216
|| |||||
Db 48895 ggcaagccaatcgaggccttccataacttctgttaggaatgcgcgaacgtctccccaggcc 48954
QY 1217 tggcgccacgcgttgcccccttaactaaagatccctgcgcgtcaggaggtgtgcgcgcggcgcg 1276
|| |||||
Db 48955 tggcgccacgcgttgcccccttaactaaagatccctgcgcgtcaggaggtgtgcgcgcggcgcg 49014
QY 1277 ctgcagctgctgggctacgcggcctgtgtactctgcggaccagcagcgtgacctcacccctg 1336
|| |||||
Db 49015 ctgcagctgctgggctacgcggcctgtgtactctgcggaccagcagcgtgacctcacccctg 49074
QY 1337 gatctggtgctgccacgagggccccagaccacttcagctggtggcctgcgcactgagaaactc 1396
|| |||||
Db 49075 gatctggtgctgccacgagggccccagaccacttcagctggtggcctgcgcactgagaaactc 49134
QY 1397 tgggccttagagcaagccccgaactgtggtgcgcaggccccaggaagcgcgactgcatggttg 1456
|| |||||
Db 49135 tgggccttagagcaagccccgaactgtggtgcgcaggccccaggaagcgcgactgcatggttg 49194
QY 1457 aaaaggagctggggcgcatggggaa-caggtccctactatatcaacccgggagtttggggctcc 1515
|| |||||
Db 49195 agaggagctggggcgcatggggaaagcaggtccctactatatcaacccgggagtttggggctcc 49254
QY 1516 tccccgtgaagtaagcaaggactgcacgtttcttctctctctctctctctctctctctctctctg 1575
|| |||||
Db 49255 tccccgtgaagtaagcaaggactgcacgtttcttctctctctctctctctctctctctctctctg 49314
QY 1576 agtcttctggagctgccttctcatcaggtgcactcttcatggaaaagcaactcttgcctcc 1635
|| |||||
Db 49315 agtcttctggagctgccttctcatcaggtgcactcttcatgg-aaagcaactcttgcctcc 49373
QY 1636 tacctcttctggcgcgaggagtaagttactgctataaattaaatgagttggtgcccaggcc 1695
|| |||||
Db 49374 tacctcttctggcgcgaggagtaagttactgctataaattaaatgagttggtgcccaggcc 49433
QY 1696 ggggtcggtggtcctgctgtaatcccagcattttgagaggctgaggggtggtgatcac 1755
|| |||||
Db 49434 ggggtcggtggtcctgctgtaatcccagcattttgagaggctgaggggtggtgatcac 49493
QY 1756 ctgagggtcaagga-ttcaaaaaccagcctggccaacacatagtgaaaacccctctctactaaaa 1814
|| |||||
Db 49494 ctgagggtcaaggagttcgaaaaccagcctggccaacacatagtgaaaacccctctctactaaaa 49553
QY 1815 atgcaaaaattagtcggcggtggtggcacactcctgtaatcccagctacttaggaggtg 1874
|| |||||
Db 49554 atgcaaaaattagtcggcggtggtggcacactcctgtaatcccagctacttaggaggtg 49613
QY 1875 aggtgggaaaaatcaccttggaactccaaagggtggaggttgcagtaagctgaaatcatgccac 1934
|| |||||
Db 49614 aggtgggagaatcaccttggaactccagaggtggaggttgcaagtaagctgagatcatgccac 49673

QY 1935 tgcaccctagcttggtggcaagcaaaactctatcaaaaaataattaataattggt 1994
|||||
Db 49674 tgcaccctagcttggtggcagagcaagactctatcaaaaaataagtaataattggt 49733
|||||
QY 1995 caaaagtccctgc 2006
|||||
Db 49734 caaaagtccctgc 49745
|||||
RESULT 5
AAD02700
ID AAD02700 standard; cDNA; 1694 BP.
XX
AC AAD02700;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.
XX
KW Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
therapy; selectin binding inhibitor; gene therapy; inflammation;
systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
asthma; hypersensitivity; rheumatic fever; tissue rejection;
chromosome 16q23.1; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
5'UTR 9..188
/*tag= a
FT 189..1376
FT /*tag= b
FT /product= "Human glycosyl sulfotransferase-4beta
(GST-4beta)"
FT /note= "CDS is referred as SEQ ID NO:21 in brief
description of the figures (page no: 4)"
FT 1377..1694
FT /*tag= c
XX
PN WO200106015-A1.
XX
PD 25-JAN-2001.
XX
PF 19-JUL-2000; 2000WO-US19741.
XX
XX 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Rosen SD, Lee JK, Hemmerich S;
XX
DR WPI; 2001-138471/14.
DR P-PSDB; AAY72640.
XX
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
diagnostic and therapeutic agent screening applications -
XX
PS Claim 6; Fig 4A; 128pp; English.
XX
CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4
beta) cDNA. GST-4 gene is found on chromosome 16q23.1.
CC
CC GST is a type 2 membrane protein useful for inhibiting a binding event
between a selectin and a selectin ligand, which comprises contacting the
selectin with a non-sulphated selectin ligand, GST and a small molecular
agent that inhibits the sulphation activity of GST. GST is also useful
in inhibiting a selectin mediated binding event. GST is useful in gene

CC therapy to treat disorders such as acute or chronic inflammation,
systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
dermatitis, myocarditis, regional enteritis, adult respiratory distress
syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
during transplantation.
XX
SQ Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;
Query Match 54.4%; Score 1111.4; DB 22; Length 1694;
Best Local Similarity 85.4%; Pred. No. 1.8e-174;
Matches 1314; Conservative 0; Mismatches 206; Indels 19; Gaps 6;
QY 140 tcgagacatctcccatggatttggatcagcgttgccagctctcccagcagccctggcgg 199
|||
Db 111 tctaatcatctcccatggatttggatcagcgttgccagctctcccagcagccctggacag 170
QY 200 tggccccagccgcccgcacgtgtggctgccacggcttctccagcaagacagtgacagtgtc 259
|||||
Db 171 tggccccagcagtcacacgtgtggctggcgcgcgtctccagcacacagtgaccgcgtc 230
QY 260 ctccctggcacagaccacctgcctcctcgtctcttcacatcatctcccggccagggccctcatcc 319
|||||
Db 231 ctccctggcgcag--acctctcctcctctcttctggtttccggccagggccctcgtcc 287
QY 320 ccagccggcggcgaggatcgtgtgcacgtgctggtgctgtctcctcgtggtggtcgcgcgc 379
|||||
Db 288 ccagcaggcggcgaggcgcgctgcatgtgtggtgctgtcctcgtggtggtcgtcgcgcgc 347
QY 380 tccttcttggccagctcttccagccagcaccgccgacgttcttaccctgatggagccgcgc 439
|||||
Db 348 tccttctggtggccaaactcttcaaccagcaccgccgacgttcttaccataatggagccgcgc 407
QY 440 tggcatgtgtggaccaccctgtcgcaggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 499
|||||
Db 408 tggcacgtgtggaccaccctgtcgcaggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 467
QY 500 gacctgatgcgtctctatcttttgtgcacatggacgtgttttgatgcctacatgccacag 559
|||||
Db 468 gacctggtgcgtccgcgtcttctcgtgcacatggacgtgtttgatgcctatctgccttgg 527
QY 560 agccgaaaacctgtccgccttttttaactgggcaacgcgcgcgcgcgcgcgcgcgcgcgcgc 619
|||||
Db 528 cgccgcaacctgtccgacctcttccagtggtggccgtgagccgtgactgtgctgcgcaccc 587
QY 620 gectgcagcgcctttcccccaggcaccatcagcaagcaggacgtatgcaagacactgtgc 679
|||||
Db 588 gcctgcagtgcctttcccccaggcgcctatcagcagcaggccgtgtgcaagccactgtgc 647
QY 680 acgcggcagccattcagcctggccccggaggcctgccgctcctacagcagcgtgtgtc 739
|||||
Db 648 gcgcggcagtccttcaacctggccccggaggcctgccgctcctacagcagcgtgtgtc 707
QY 740 aaggaggtgcgtcttcaacctgcaggtgcttaccgccgtgctcagcagcccccgcgc 799
|||||
Db 708 aaggaggtgcgtcttcaacctgcaggtgcttaccgccgtgctcagcagcccccgcgc 767
QY 800 aacctgcgcacgtgcacctggtgcgcgcaccgccggccgctgctgcctcccggaggcg 859
|||||
Db 768 aacctacgcacgtgcacctggtgcgcgcaccgccggccgctgctgcctcccggaggcag 827
QY 860 gcgggccccgatactggcacgcgcacaacggcatcgtgctgggcaccacaacggcaagtgggtg 919
|||||
Db 828 acagccaaaggctctgtggcgttgacaacggcatcgtgctgggcaccaacggcagcgtgggtg 887
QY 920 gaggccgaccctcaacctgcgcctgattcgcaggtgtgcgcagccacagtcgcgcacatgcc 979
|||||
Db 888 gaggccgacccccgcctgcgcgtggtgcgcgcaggtgtgccctagccacagtcacgcacatgcc 947

QY 980 gagggcgccacactcaagccgcaccccttctgcccgcgtaccgcctggtgcgcttc 1039
|||||
Db 948 gagggcgccacactcaagccgcaccccttctgcccgcgtaccgcctggtgcgcttc 1007
QY 1040 gaggaacctggcgggagccgctggcagagatccgcgcactctacgccttcacccgcctg 1099
|||||
Db 1008 gaggaacctggcgggagccgctggcagaaatccgtgcgtctacgccttcactggctc 1067
QY 1100 acctcacgcacagctcgaggccctggatccacaacatcacccacgggtcggggatcggc 1159
|||||
Db 1068 agtctcacgccacagctcgaggccctggatccataacatcacccacggatctggacctggt 1127
QY 1160 aagccaatcgaggcccttcataacttctgtctaggaatgcgcgaacgtctcccaggcctgg 1219
|||||
Db 1128 gcgcgcgcgaagccttcaagacttctgccaggaatgcgtcaacgtctcccaggcctgg 1187
QY 1220 cgccacgcgttgcccttcaactaagatccctgcgcgtcaggagggtgtgcgcgcgcgctg 1279
|||||
Db 1188 cgccatgcgtgcccttgcgaagatccgcgcgtcaggaaactgtgcgtggtgcgctg 1247
QY 1280 cagctgctgggctacggccctgtgtactctgcggacacgacgctgacctcacccctggat 1339
|||||
Db 1248 cagctgctgggctacggccctgtgtactctgaggacgagcagcgcaacctgcacctggt 1307
QY 1340 ctggtgctgccacagggccacagaccacttcagctgggcatc--gcctgactgagaaactct 1397
|||||
Db 1308 ctggtgctgccacagggcctgaacggttcaacttgcgtggaatccaccgcctgcacccc 1367
QY 1398 gggccttagagcaagcccc--gaactgtggtgcgcaggccacgagcagctgacctcacccctggtg 1455
|||||
Db 1368 cgaatagtgaggccacagttgtagcaggcgctagggcccgaggagagagtgcatggtg 1427
QY 1456 gaaaaggagctgggcgcatgggaa-caggtccctactatcaacccgggagtttggggt- 1513
|||||
Db 1428 cagaggggctgggcgccacggagagcaggtccctatatattgaccaaaggagtttggta 1487
QY 1514 -----cctccctgaagtaagcaaggactgcacgttttcttctctctgattctcg 1564
|||||
Db 1488 cgacccctccctcccaagtaggcaaggactgcacgtttcttctctctgattctcg 1547
QY 1565 gtttctcttgagttcttctggagctgaccttcctcatcaggtgcactcttcattgaaagca 1624
|||||
Db 1548 gtttctcttgagttcttctggagctgaccttcctcatcaggtgcactcttcattgaaagca 1606
QY 1625 actcttgccctacctcttcttctggcgagcagggagtaagt 1663
|||||
Db 1607 actcttgccctgctctctctctggcgacaggggtgctgctt 1645

RESULT 6
AAD24670
ID AAD24670 standard; cDNA; 1647 BP.

AC AAD24670;

DT 12-MAR-2002 (first entry)

DE Human drug metabolising enzyme (DME)-5 cDNA.

XX Human; drug metabolising enzyme; gene therapy; autoimmune disorder;
KW inflammatory disorder; acquired immune deficiency syndrome; infection;
KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
KW cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;
KW gastrointestinal disorder; metabolic disorder; developmental disorder;
KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
KW DME-5; ss.

XX Homo sapiens.

OS Homo sapiens.

XX Homo sapiens.

FH Key Location/Qualifiers

CDS 1..1188
/*tag= a
/product= "Human DME-5 protein"
1..96
sig_peptide
/*tag= b
mat_peptide 97..1185
/*tag= c
/*tag= c
/product= "Human mature DME-5 protein #1"
1..105
sig_peptide /*tag= d
mat_peptide 106..1185
/*tag= e
/*tag= e
/product= "Human mature DME-5 protein #2"

WO200179468-A2.

PD 25-OCT-2001.

PF 12-APR-2001; 2001WO-US11869.

PR 13-APR-2000; 2000US-197590P.

PR 19-APR-2000; 2000US-198403P.

PR 28-APR-2000; 2000US-200185P.

PR 05-MAY-2000; 2000US-202234P.

PR 11-MAY-2000; 2000US-203509P.

XX (INCY-) INCYTE GENOMICS INC.

XX Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;

PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;

PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;

PI Au-Young J;

XX WPI: 2002-066363/09.

DR P-PSDB; AAE15438.

DR P-PSDB; AAE15438.

XX Novel isolated human drug metabolizing enzymes referred as DME 1-10,

PT useful for diagnosing, treating, or preventing disorders associated

PT with aberrant expression of DME such as allergy, anemia, asthma,

PT infertility -

XX Claim 5; Page 139; 143pp; English.

XX The invention relates to human drug metabolising enzymes referred as

CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides

CC of the invention are useful for assessing toxicity of test compounds

CC and in gene therapy. Sequences of the invention are useful in the

CC diagnosis, prevention and treatment of autoimmune/inflammatory

CC disorders such as acquired immune deficiency syndrome (AIDS), adult

CC asthma, autoimmune haemolytic anaemia, allergies, anaemia, atherosclerosis,

CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's

CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple

CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus

CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,

CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic

CC infections; cell proliferative disorders such as actinic keratosis,

CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,

CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,

CC epilepsy; endocrine disorders such as disorders of the hypothalamus

CC and pituitary resulting from lesions such as primary brain tumours,

CC adenomas, infarction associated with pregnancy, aneurysms, vascular

CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,

CC glaucoma, pigmentosa; metabolic disorders such as Addison's disease,

CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,

CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,

CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such

CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,

CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's

CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental

CC disorders. The present sequence is human DME-5 protein cDNA.

XX Sequence 1647 BP; 287 A; 560 C; 501 G; 299 T; 0 other;

probe. The new human transferases and polynucleotides can be used in the diagnosis, prevention and treatment (including gene therapy and antisense therapy) of cancer, developmental disorders, gastrointestinal disorders, genetic disorders, immunological disorders, neurological disorders, reproductive disorders, and smooth muscle disorders.

Query Match 18.4%; Score 377; DB 21; Length 2065;
Best Local Similarity 63.4%; Pred. No. 1.5e-53;
Matches 644: Conservative 0; Mismatches 360; Indels 11; Gaps 4;

QY	335	gacgtgtgcacgtcgtggctgtgtcctcgctggcgtcggtccatccttcttggtgccag	394
Dd	294	gagcgcatagcacgtgctgttctgtcttcctggcgtctgttctttcttgttgggacag	353
QY	395	ctcttcaggccagcacccgaagtcttttacactgatggagccccgtggcaatgtgtggacc	454
Dd	354	ctttttggcagcaccccagaigtcttctacctgatggagccccgtggcaagtgtggatg	413
QY	455	acctgtcgcagggcagcgcggaacgctgcacatggccgtgcggaacctgatgcgtct	514
Dd	414	accttcaagcagacccgcctggatgctgcacatggctgtcgggatctgatacgggcc	473
QY	515	atatttttgtgcacatggacgtgttgtgatgcctacatgccacagag--ccgaaaacctg	571
Dd	474	gtcttttgtgcacatgagcgtctttgatgcctacatggaaacctggtccccggagacag	533
QY	572	tccgcctttttcaaactggcaacgagccgcgcgtgtgtctgcgccgccgcctgcagcgcc	631
Dd	534	tccagcctcttcagtgggagaacagccggctactgcaggctcctgtgtctgcacctgcctgtgacatc	593
QY	632	tttccccgaggcaccatcagcaagcaggacgtatgcaagacactgtgcacgcggcag-cc	690
Dd	594	-atcccacaagatgaatcatcccccggtcactgcaggctcctgtgcagtccaacagccc	652
QY	691	attcagcctggccccggaggcctgcctcctacagccacgtgtgtgtctcaaggagtgcg	750
Dd	653	tttgaaagtgttgagaaaacctgccgtcctacagccacgtgtgtctcaaggagtgcg	712
QY	751	ctttctcaacctgcaggtgctctacccgctctcagcgacccccgcgtcaacctgcgcac	810
Dd	713	ctttctcaacctgcagtcctctaccccgtctgtaagagacccccctcctcaacctgcacat	772
QY	811	cgtgaacctggtgcgcgaccccgggccgtgtgtgcgtctccgggagggcgggcccgat	870
Dd	773	cgtgaacctggtccgggacccccgggcgtgttccgttccccgagaacgcacaaaaggaga	832
QY	871	actggcacgcgacaacggcatcgtgtgtgggaaccaoggaagtgggtggagccgaccc	930
Dd	833	tcctcatgttgacagtgcattgtgatggggcagcatgagcagaaactcaagaaggagga	892
QY	931	tcacctgcgcctgattcgcgaggtgtgcgcgacgcacgtgcgcacgtccgagggccgccac	990
Dd	893	ccaacctactatgtgatgcaggtcatctgccaaagccagctggagatctacaagacct	952
QY	991	actcaagccgccacccttcctgcgcggccgcctaccgcctgggtgcgttcgaggaacctggc	1050
Dd	953	ccagtccttgccaaagccctgcaggaacgcctacctgcttggcgctatgaggaacctggc	1012
QY	1051	gcgggagccgtggcagagatccgcgcactctacgccttcacccggcctgacccctcacgcc	1110
Dd	1013	tcgagccccgtggcccagaccttcccgaatgtatgaattcgtggattgggaattcttgcc	1072
QY	1111	acagctcgaggcctggatccacaacatcacccacgggtcggggatcggcaagccaatcga	1170
Dd	1073	ccatcttcagacctgggtgcataaatcacccagggcaaggggcatgggtga-----cca	1126
QY	1171	ggccttccatacttcgtctaggaatgcgcgcaacgtctccccaggcctggcgccacgcgtt	1230
Dd	1127	ccgrrtccacacaaatgccaggaatgcccttaatgctccccaggctggcgctgggtcttt	1185

Db 648 accagcagcacagcctggaagctgcacatggtgtgcggatcttctgcgttcogtcttc 707
QY 521 ttgtgcgacatggacgtgttttgatgcctacatg---ccacagagccgaaacctgtccgcc 577
Db 708 ctgtgtgacatgagcgtcttttgatgcctacatgaaccaggcccccggaaacagtccagc 767
QY 578 tttttcaactgggcaacagagccgcgctgtgtctgcgcgcgcgcctgcagcgcctttccc 637
Db 768 ctcttcagtgaggcaaaagccgggcccctgtgtctcagcgcctgtgtgtgactttctccct 827
QY 638 cgaggccaccatcagcaagcaggacgtatgcaagacactgtgcacgcgccagccattcagc 697
Db 828 gccacagatcagctcacccaagcactgcaagctgtctgtgcgtgcagcagcccttgat 887
QY 698 ctggcccgaggccctgcgcctcctacagccacgtggtgtctcaaggaggtgcgtctcttc 757
Db 888 atggtggagaaggcctgcgctctcagcgcttcgtgtgtactcaaggaggtgcgtttctc 947
QY 758 aacctgcaggtgtctacccgctgctcagcagaccgcgcgtcaacctgcgcacatcgtgcac 817
Db 948 agcctgcaggccctctatccactactcagggacccttcctcctcaacctgcacgtcgtgcac 1007
QY 818 ctggtgcgcgacccgcggccgtgctgcgctcccgaggagcgggcccgatactggca 877
Db 1008 ctggtccgagaccccgggccgtgttcgcgtatcccgaggagcacaccacatagaactcatg 1067
QY 878 cgcacaacggcatctgtgtggggcaccacaacggcaagtgggtggaggccgacccctaacctg 937
Db 1068 gttgacagtcatattgtgtagggcagcatttggaaaacgatcaaggaggagaaccagccc 1127
QY 938 cgcctgattcgcgaggtgtgcccagcccacgtgcgcctgcgcagcccgagccgccaactcaag 997
Db 1128 tattatgccatgaagatcatctgcaaaaggccaggtggacatagtaaggccatcaaac 1187
QY 998 ccgcacaccttctcgcgggcgcgtaccgcctggtgcgcttcgaggaacctggcgggag 1057
Db 1188 ctccctgaagctctcagcagcgtacctgttctgaggtatgaggacctggttcgggca 1247
QY 1058 ccgctggcagagatccgcgcactctacgccttcacccggcctgacctcagccacacagctc 1117
Db 1248 cccctggcccagagaccagactatataaatttgggttggttgatttttggccccacctc 1307
QY 1118 gaggcctggatccacaacatcacccacgggtcggggatcggcaagccaatcgaggcccttc 1177
Db 1308 caaacatgggtttacaatgtcacccgcggcaaggcatgggtcagc-----atgccttc 1361
QY 1178 catacttcgtctaggaatgcgcgcaacgtctcccaggccctggcgcacacgcttgcccttc 1237
Db 1362 catactaacgcagggaacgcctcctcaacgtctcagggcgtggcgttggttccttaccttac 1421
QY 1238 actaagatcctgcgctgcaggaggtgtgtggccgcgcgcgtgcagctgcgtgggtaccgg 1297
Db 1422 gaaaaggtttcccagcttcaagatgcctgcggtgaggtatggatttgcgtgggatacctc 1481
QY 1298 cctgtgtactctcgggaccagcagcgtgaacctcaacctggatctggtg 1345
Db 1482 caggtcagatctcaacaagaacaaggcaacctgtccctggatcttctg 1529

RESULT 14
AAK91803
ID AAK91803 standard; cDNA; 877 BP.
XX
AC AAK91803;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA 5'-end sequence, SEQ ID NO: 263.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.

XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 2; SEQ ID NO 263; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 14.6%; Score 297.6; DB 22; Length 877;
Best Local Similarity 71.4%; Pred. No. 1.7e-40;
Matches 406; Conservative 0; Mismatches 160; Indels 3; Gaps 1;

QY 335 gatcgtgtgcacgtgctgtgtctcctcgtggcgtcggtcggtcctatccttcttgggccag 394
Db 189 gagcgcgatgcacgtgctgtctctcctcgtgcgtcggtccttcttcttggggcag 248
QY 395 ctcttcagccagcaccgcagcgtcttctacctgatgagcccgctggcatgtgtggacc 454
Db 249 cttttgggcagcaccagatgtttctacctgatggagcccgctggcacgtgtgatg 308
QY 455 accctgtcgcagggcagcgcggcaacgctgcacatggccgtgcgcacctgatcgctct 514
Db 309 accttcaagcagagcaccgcctggatgctgcacatggctgtgcgggatctgatacggcc 368
QY 515 atctttttgtgcgacatggacgtgttttgatgcctaatgccacagag---ccgaaacctg 571
Db 369 gtcttcttgcgacatgagcgtcttctgatgcctacatggaacctggtcccgagacag 428
QY 572 tcgcctttttcaactgggcaacagcagccgcgcgtgtgtcgcgcgcgcctgcagcgc 631
Db 429 tcagcctctttcagtgaggagacagccggccctgtgttctgcacctgcctgtgacatc 488
QY 632 ttcccccgaggccaccatcagcaagcaggacgtatgcaagacactgtgcacgcggcagcca 691
Db 489 atccccacaagatgaaatcatcccccggtcactgcaggtcctgtgcagtcagtcaacagccc 548
QY 692 ttcagcctggcccgaggcctgcgcctcctacagccacgtgtgtgtctaaaggaggtgcgc 751
Db 549 tttgaggtggtgagaaggcctgcgcctcctacagccacgtgtgtgtctaaaggaggtgcgc 608
QY 752 ttcttcaacctgcaggtgtcttaccgcgtgtctcagcgaccccgcgctcaacctgcgcac 811

Db 609 ttotccaacctgcagtcctctatccccgctgctgaaagacccccctccctcaacctgcatac 668
QY 812 gtgaacctggtgcgcagaccgcggcgccgtgctgcgtcccgaggcgggcccgata 871
Db 669 gtgcacctggtccgggaccccgggccgtgttcctcccgagaacgcacaaaggagat 728
QY 872 ctggcacgcgacaaacggcatcgtgctggg 900
Db 729 ctcatgattgacaatcgcatgngatggg 757

RESULT 15
AAK93921
ID AAK93921 standard; cDNA; 877 BP.
XX
AC AAK93921;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA clone representative sequence, SEQ ID NO: 2381.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Example 11; SEQ ID NO 2381; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 14.6%; Score 297.6; DB 22; Length 877;
Best Local Similarity 71.4%; Pred. No. 1.7e-40;
Matches 406; Conservative 0; Mismatches 160; Indels 3; Gaps 1;
QY 335 gatcgtgtgcacgtgctggtgctgtcctcgtggcgctcgggtcatccttttgggccag 394
Db 189 gagcgcacgtgcgtggttctgtctcctggcgctcgtggtctcttttggggcag 248
QY 395 ctcttcagccagacccccgacgtcttctacctgatggagccccgctggtgtggacc 454

Db 249 ctttttggcagcaccagatgtttctacctgatggagcccgctgcacgtgtggatg 308
QY 455 acctgtgcagggcagcgcggaacgctgcacatggccgtgcgcgacotgatgcgtct 514
Db 309 accttcagcagagcaccgcctggatgctgcacatggctgtcggggtatgatacggcc 368
QY 515 atctttttgtgcgacatggacgtgttttgatgcctacatgccacagag--ccgaaacctg 571
Db 369 gtcttcttgtgcgacatgagcgtcttttgatgcctacatggaaacctgtccccggagacag 428
QY 572 tcggcctttttcaacctgggcaacgagcgcgcgctgtgtctgcgcgcgcctgcagcgcc 631
Db 429 tcagcctcttccagtgggagaaacagccgggccccctgtgttctgcacctgcctgtgacatc 488
QY 632 tttoooooagggcaccatcagcaagcagcagctatgcaagacactgtgcacgcggcagcca 691
Db 489 atccccacaagatgaaatcatccccgggggtcactgcaggctcctgtgcagtcacacagccc 548
QY 692 ttcagcctggccccgggaggcctgcgcgtccgtcctacagccacgtgtgtctcaaggaggtgcgc 751
Db 549 tttagagtggtggagaaaggcctgcgcgtcctcctacagccacgtgtgtctcaaggaggtgcgc 608
QY 752 ttcttcaacctgcaggtgctctacccgctgtcctacagcagccccgcgtcaacctgcgcac 811
Db 609 ttcttcaacctgcagtcctcctacccgctgtgtaaaagacccccctccctcaacctgcatac 668
QY 812 gtgcacctggtgcgcgacccccggcgcgctgtgtcgtccccgggaggcgggcccgata 871
Db 669 gtgcacctggtccgggacccccggcgcgctgttccgttccccgagaaacgacaaaaggagat 728
QY 872 ctggcacgcgacaaacggcatcgtgctggg 900
Db 729 ctcatgattgacaatcgcatgngatggg 757

Search completed: September 24, 2002, 19:04:23
Job time: 45322 sec


```

Db 662 tttaggtggtggaagccctgccgtccctacagccacgtggtgctcaaggagtgcg 721
QY 535 ttcttaacctgcaggtgctctacccgctgctcagcgaccccgctcaacctgcgcac 594
Db 722 ttcttaacctgcagtcctctacccgctgctgaaagacccctccctcaacctgcatac 781
QY 595 gtgcacctggtgcgagaccccgccggtgctgcgctcccgaggcgccgcccgcgata 654
Db 782 gtgcacctggtgcgagaccccgccggtgctgcgctcccgaggcgccgcccgcgata 841
QY 655 ctggcacgcgacaacggcatctgctgggcaaccaacggcaagtgggtggagccgaccc 714
Db 842 ctcatgattgacagtgcattgtgatgggcagcatgagcagaactcaaggaggagac 901
QY 715 caactgcgcctgattgcgaggtgtgcgcgcagccacgtgcgcacgcggcgccaca 774
Db 902 caacctactatgtgaggtcaggtcatctgccaaagccagctggagatctacaagacc 961
QY 775 ctcaagccgcaccccttccctgcgcgcgcctaccgctgctggtgcgcttogaggacctggcg 834
Db 962 cagtccttgcccaagccctgcaggaacgctacacctgtgtgcgtatgaggacctggct 1021
QY 835 cgggagccgctggcagagatccgcgcgcactctacgccttcacccgacctgacctcaaggcca 894
Db 1022 cgaagccctgtggccagacttcccgaaatgtatgaattcgtggattggaattcttgccc 1081
QY 895 cagctcgagccctggatccacaacatcacccacgggtggggatcggcgaagccaatcgag 954
Db 1082 catctcagacctgggtgcataacatcacccgagggcaaggccatgggtga-----ccac 1135
QY 955 gccttcactactctctaggaatgcgcgaacgtctccagacctgcccagccagcgttg 1014
Db 1136 gcttccacacaaatgccaggatgcccttaattgtctccagcgttgccgctggtcttg 1195
QY 1015 ccttcaactaatcctgcgcgtgcagaggtgtgcgcgcgcgcgtgcagctgctgggc 1074
Db 1196 ccctatgaaaaggtttctcgacttcagaaagcctgtggcgatgccatgaatttgcctgggc 1255
QY 1075 taccggcctgtgtactctgcggaccagcagctgacctcaacctggatctgggtg 1128
Db 1256 taccgccaagtgcagatctgaacaaagacagaaaacctgttgctggatctctcg 1309
```

```

RESULT 2
US-08-899-514-1
; Sequence 1, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
; APPLICANT: HABUCHI, OSAMI
; APPLICANT: FUKUTA, MASAKAZU
; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
; TITLE OF INVENTION: FOR THE POLYPEPTIDE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,514
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; 
```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DANIEL E ALTMAN
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714 760 0404
; TELEFAX: 714 760 9502
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2156
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Human
; TISSUE TYPE: Fetal brain
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 147..1583
; IDENTIFICATION METHOD: S
US-08-899-514-1
```

```

Query Match 20.1%; Score 235.2; DB 2; Length 2156;
Best Local Similarity 54.0%; Pred. No. 8.8e-41;
Matches 589; Conservative 0; Mismatches 468; Indels 33; Gaps 4;

QY 91 gggccctcatcccccagcggcgaggatcgtgtgcacgtgctgctcctcgtg 150
Db 510 GAGCCGCCAGACCGGCGGTGGCGGGCCCCCGGCCACGTGCTGCTCATGGCCACCAG 569
QY 151 cgctcgggctcatccttcttggccagctcttcagccagcaccgcagcgtcttctac 210
Db 570 CGCACCGGCTCCTCGTTCGTGGCGAGTCTTCAACCCAGCAGGGCAACATCTTCTAC 629
QY 211 atggagcccgctggcagctgtgtggaccacctgtc-----gcagggcagcgcg 258
Db 630 TTCGAGCCGCTGTGGCACATCGAGCGGCACAGTGTCTTTCGAGCCGGGGCGCCAAC 689
QY 259 gcaacgctgcacatggcgtgcgcgacctgatgcgtctatctttttgtgcacatggac 318
Db 690 GCGGGCTCGGCCCTGGTGTACCGCGACGCTCTTCCCTGTCGACCTCTTCCCTGTCGAC 749
QY 319 gtgttgatgcctaca-----tgcaacagagccgaaacctgtccgctttttcaac 369
Db 750 GTGCTGGAGCACTTCATACGCCCGTGCCTGAGGACCCACCTGACTGACTGCTCCGC 809
QY 370 tgggcaacagcgcgcgtgtgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 429
Db 810 CGGGGCTCCAGCGCTCCCTGTGCGAGGACCCGCTGTGTACGCCCTTCGTCAAGAAGGTC 869
QY 430 atcagcaagcagcagctatgcaagacactgtgcacgcggcagccattcagcctggccc 489
Db 870 TTCGAGAAGTACCACTGCAAGAACCCGCGCTGCGGCCCTCCCTCAACGTGACGCTGG 929
QY 490 gaggcctgccgctcctacagccacgtggtgctcaaggaggtgcgcttcttcaacctgcag 549
Db 930 GAGGCTGCCCGCCGAAGGACACATGGCCCTCAAGCGGTGCGCATCCGCGAGCTGGAG 989
QY 550 gtgctctacccgctgcagcagcccccgcgtcaacctgcgcagctgtgcacctggtgcgc 609
Db 990 TTCCTGCAGCGCTGGCCGAGGACCCCGCCCTGGACCTGCGCGTTCATCCAGCTGGTG 1049
QY 610 gacccgcggcgcgtgctgcgctcccgaggagcgggggcccgatactggcacgcgacaac 669
Db 1050 GACCCCGGGCGCTGCTGGCCTCGCGCATGGTGGCCTTCGCCGGCAAGTATAAGACCT 1109
QY 670 ggcacgtgctgggcacacaacggcaagtgggtggagggcgacccctcacctgcgcctg 729
Db 1110 AAGAAGTGGCTGGACGACGAGGGCCAGGACGGCCTGAGGGAAGAGGAGGTGCAGCGG 1169
```



```

;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
;
; US-08-804-227C-1

Query Match          5.5%; Score 64.6; DB 2; Length 43280;
Best Local Similarity 44.7%; Pred. No. 4.8e-05;
Matches 305; Conservative 0; Mismatches 369; Indels 9; Gaps 1;

QY 396 gccgcccgcctgcagcgccctttccccgagggcacccatcagcaagcaggacgtatgcaagac 455
      ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 29720 GCCGCTCACCGCCTACGGCATACGGCATACGGCAGCGCCCGCACCGCCTTGCGCACCCCTCAGCCAGGC 29779

QY 456 actgtgcacgcggcgagccattcagccttgcccgaggccctgcgcgtcctctacagccacgt 515
      ||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 29780 CCGGCACACCGCGCAAGCTCGTGTGACGGTGCCTGCCGGATTTCGACACCCACCGCACGGT 29839

QY 516 ggtgctcaagaggtgcgctttcttcaacctgcaggtgctctacccgcgtcgtcagcgaccc 575
      ||| |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 29840 GCTCCTCACCGCGCGCACGGGCACGGCTCGGCGCAGACACTCGCCCGCCATCTGGTCAACC- 29898

QY 576 cgcgctcaacctgcgcacgtgtgcacctggtgcgcgacccgcggccgtgtgcgctcccg 635
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29899 -----GCCACGGCGGTACGGCACCTGCTGCTCGCCGCCGACACCGGGCGCGGCCCGA 29950

QY 636 ggaggcgggcgcccgatactggcacgcgcgaacggcgcgtgtgtgggcacaaacygcaa 695
      ||| |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 29951 GGGCGTCGCGGAAGTGAATGGTGAATGGGAGTGGCGCGAGTGGCGCGCGAGTCCGGTCCGCGC 30010

QY 696 gtgggtggaggccacctcacctgcgcctgattgcgaggtgtgtccgcagcgcacgtgcg 755
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30011 CTGGACGCGCGCGGACCGGACGGGCTCACCGAACTCTCGCCGGAATCCCGTCGAGCA 30070

QY 756 catcgccgaggccgcacactcaagccgcaccttctctgcgcggccgtacgcgcctggt 815
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30071 CCCGCTCGGCGCGCTCTCCACCGCCGCGGCACCTTCGACGACGCGCACCATCCCTCACT 30130

QY 816 gcgcttcgaggacctgcgcgcggagccgctggcagagatccgcgcactctacgccttcac 875
      || | |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30131 GACCGCGGAGAACATCGACAACGTGCTGCGGCCCAAGGCCGACGCCGTGCTCAACCTGCA 30190

QY 876 cggcctgacctcacgcacagctcgcagccgtgagccctggatccacaacatcacccgggtcg 935
      || |||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30191 CGAGCTGACCCCGGACGCGACCTCTCGGCGTTCGTCCTCTACTCTCTCTCTCTCTCTCT 30250

QY 936 gatcggaagccaatcagggcccttcacattcgttaggaatgcgcgcaacgtctccca 995
      || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30251 GCTCGGCAGCCCGCGCCAGGCGCGCCTACGCGCGCGCAACGCCTTCCTGGACGGCTTCGC 30310

QY 996 ggcctggccacgcgttgccttcactaagatcctgcgcgtgcagaggtgtgcgcgg 1055
      | |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30311 CCGATACCGCAAGGGCCTCGGCTGCGCGGCTCTCGCTGGCTGGGACTGTGGGGCAG 30370

QY 1056 cgcgctgcagctgctgggtacc 1078
      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30371 CAACAGCCGATGGCGGGGCCACC 30393

RESULT 7
US-09-040-984-63
; Sequence 63, Application US/09040984
; Patent No. 6210883
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, TongTong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
```

```

;
; TITLE OF INVENTION: OF LUNG CANCER
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,984
; FILING DATE: 18-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.456
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-282-6031
; TELEX:
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 731 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-040-984-63

Query Match          4.9%; Score 57.8; DB 4; Length 731;
Best Local Similarity 58.4%; Pred. No. 0.00068;
Matches 101; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 515 tgggtgctcaaggaggtgcgtcttctcaacctgcaggtgctctaccgcgtgcgcagcacc 574
      ||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 3 TAGTCATAAAGGGTGTGCGGCTCTTCGACGTGCGGCTCTTGGCGCCACTGCTGCGAGACC 62

QY 575 ccgcgctcaacctgcgcacatcgtgcacctggtgcgcgaccccggggcgtgctgcgcctccc 634
      ||| ||| |||| | ||| ||| |||| | ||| ||| |||| | ||| ||| |||| | ||| |||
Db 63 CGGCCCTGGACCTCAAGGTCATCCACTTGGTGGTGATCCCCCGCGGTGGCGAGTTTCC 122

QY 635 gggaggcgcgggcccgatctctggcacgcgcgacaacgcgcacgtgctgggcacc 687
      || |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 123 GGATCCGCTCGCGCCACGGCCCTCATCCGTGAGAGCCTACAGGTGGTGGCAGC 175

RESULT 8
US-09-123-912-63
; Sequence 63, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C1
; CURRENT APPLICATION NUMBER: US/09/123,912A
; CURRENT FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/040,802
; PRIOR FILING DATE: 1998-03-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 19:58:31 ; Search time 13.53 Seconds
(without alignments)
1116.085 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRFSSKTVTVLLLAQTT.....LTLDLVLPRGPDHFSWSPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	540.5	26.4	458	1 C6ST_CHICK	Q92179 gallus gall
2	105	5.1	359	1 TPSE_CAEEL	Q20351 caenorhabdi
3	102.5	5.0	388	1 YM67_MYCTU	Q50695 mycobacteri
4	96	4.7	388	1 ODPT_HUMAN	P29803 homo sapien
5	96	4.7	707	1 PAL1_IPOBA	P14166 ipomoea bat
6	95	4.6	885	1 CLPC_ODOSI	P49574 odontella s
7	94.5	4.6	1581	1 VGLP_BEV	P23052 berne virus
8	90.5	4.4	380	1 TPSE_CAEEL	O77081 caenorhabdi
9	89	4.3	371	1 CYB_LATCO	Q9mlj4 laticauda c
10	88.5	4.3	249	1 NODH_RHTR	P52994 rhizobium t
11	86.5	4.2	376	1 TPS2_MOUSE	O88856 mus musculu
12	86.5	4.2	377	1 TPS2_HUMAN	O60704 homo sapien
13	86.5	4.2	1608	1 RRPL_BDV	P52639 borna disea
14	85.5	4.2	221	1 YPRB_CORGL	P46584 corynebacte
15	85.5	4.2	389	1 ODPA_PIG	P29804 sus scrofa
16	85	4.1	1182	1 RTP2_TRYBG	P15594 trypanosoma
17	84.5	4.1	390	1 ODPA_MOUSE	P35486 mus musculu
18	84.5	4.1	390	1 ODPA_RAT	P26284 rattus norv
19	84.5	4.1	491	1 C2FI_HUMAN	P24903 homo sapien
20	84.5	4.1	1844	1 POLR_TYMA	P20128 turnip yell
21	83.5	4.1	390	1 ODPA_HUMAN	P08559 homo sapien
22	83.5	4.1	1844	1 POLR_TYMV	P10358 turnip yell
23	83.5	4.1	1844	1 POLR_TYMC	P28477 turnip yell
24	83	4.0	363	1 ODPA_SMIMA	P52900 sminthopsis
25	83	4.0	391	1 ODPT_MOUSE	P35487 mus musculu
26	82.5	4.0	1065	1 EMBB_MYCAV	P71486 mycobacteri
27	82	4.0	746	1 PPK_STRCO	Q9kzv6 streptomyce
28	81.5	4.0	703	1 PPK_STRGR	Q9eus8 streptomyce
29	81.5	4.0	715	1 PERE_HUMAN	P11678 homo sapien
30	81.5	4.0	3011	1 POLG_HCV1	P26664 h genome po
31	81	3.9	882	1 Y890_MYCTU	Q10550 mycobacteri
32	81	3.9	972	1 MTRA_YEAST	Q99189 saccharomyc
33	80.5	3.9	704	1 DP3E_RHOCA	O68045 rhodobacter

34	80.5	3.9	976	1	EPAL_HUMAN	P21709 homo sapien
35	80.5	3.9	1942	1	Y054_HUMAN	P42694 homo sapien
36	80	3.9	247	1	NOH4_RHIME	P06237 rhizobium m
37	80	3.9	328	1	RPOA_NEIMA	Q9jr06 neisseria m
38	80	3.9	392	1	BM15_HUMAN	O95972 homo sapien
39	80	3.9	415	1	CXA3_RAT	P29414 rattus norv
40	80	3.9	416	1	CXA3_MOUSE	Q64448 mus musculu
41	80	3.9	933	1	OD01_ECOLI	P07015 escherichia
42	79.5	3.9	559	1	MASY_MAIZE	P49081 zea mays (m
43	79	3.9	738	1	PLAK_XENLA	P30998 xenopus lae
44	78	3.8	462	1	TRPE_THETH	P05378 thermus aqu
45	77.5	3.8	370	1	TPS1_HUMAN	O60507 homo sapien

ALIGNMENTS

RESULT 1
C6ST_CHICK
ID C6ST_CHICK STANDARD; PRT; 458 AA.
AC Q92179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Chondroitin 6-sulfotransferase (EC 2.8.2.17) (C6ST).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.
RC STRAIN=WHITE LEGHORN; TISSUE=Embryonic chondrocytes;
RX MEDLINE=95355490; PubMed=7629189;
RA Fukuta M., Uchimura K., Nakashima K., Kato M., Kimata K.,
Shinomura T., Habuchi O.;
RT "Molecular cloning and expression of chick chondrocyte chondroitin 6-
sulfotransferase.";
RT J. Biol. Chem. 270:18575-18580(1995).
RL -!- FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN
SULFATE.
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + chondroitin =
adenosine 3',5'-bisphosphate + chondroitin 6'-sulfate.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE
(BY SIMILARITY).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D49915; BAA08655.1; ..
DR Transferase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.
KW DOMAIN 1 23 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 24 37 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 38 458 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 458 AA; 52253 MW; C9A3B7D0A5086F0C CRC64;

Query Match 26.4%; Score 540.5; DB 1; Length 458;
Best Local Similarity 34.3%; Pred. No. 2.2e-41;
Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;
QY 36 AGGEDRVHVLVLSSWRSGSSFLGQLFSQHPDVFYLMPEAHV--WTTLSQGSATLHMAY 93

```
Db 107 AAEPRRHVLIMATRTGSSFVGEFFNQGNIFYLEPDLWHIERTVTFEPGGANAVGSAL 166
QY 94 --RDLMSIFLCDMDVFDAYM---PQSRNLSAFFNWATSRALCSPACSAFPRGTISKQD 148
Db 167 VYRDVLOQLLCDLIYLESFISPAPAEHLTAALFRRGSSSHSCEEPVCTPSLKKVFEKYH 226
QY 149 VCKTLCTROPFSLAREACRSYSHVVLKEVFFNLQVLYPLLSDPALNLRIVHLVRDPRAV 208
Db 227 CKNRRCGPLNITLAAEACRRKQHMALKTVIRIQLEFLQPLAEDPRDLRIIQLVRDPRAV 286
QY 209 LRSREAAGPILARDNGIVLGTNGKWVE-----ADPHLRILIREVCRSHVRIAEAA 259
Db 287 LVSRMVA-----FSGKYESWKWAAEAGEAPLOEDEVOQLRGNCES-IRLSAEI 335
QY 260 KPPPFRLGRYRLVRFEDLAREPLAEIRALYAFTGLTLTPQLEAWTHNTHGSGIKPIEA 319
Db 336 RQPRWLGRYMLVRYEDVARAPLRKALEMYRFAGIHPTPQVEEWIRANTQAP---QDS 392
QY 320 FHTSSRNARNVSQAWRHALPFTKILRVQEVCGALQLLGYRVPVYSADOQRDLTL 379
Db 393 IYSTQKNSSEHQFEKWRFSIPFKLAQVQDACEPAMRLFGYKLASSAQELTNRSL 451
QY 380 GP 381
Db 452 GP 453

RESULT 2
TPSB_CAEEL STANDARD; PRT; 359 AA.
AC Q20351;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative protein-tyrosine sulfotransferase (EC 2.8.2.20)
DE (Tyrosyl)protein sulfotransferase (TPST).
GN F42G9.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Taich A.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN
CC ACIDIC MOTIFS OF POLYPEPTIDES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + protein tyrosine =
CC adenosine 3',5'-bisphosphate + protein tyrosine-O-sulfate.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi membrane (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00051; AAA91354.1; -.
CC WormPep; F42G9.8; CE07235.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
KW Hypothetical protein; Transferase; Transmembrane; Glycoprotein;
KW Signal-anchor. 1 27 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 28 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT TRANSMEM 28 48 (POTENTIAL).
FT FT 49 359 LUMENAL, CATALYTIC (POTENTIAL).
FT FT
```

```
SQ SEQUENCE 359 AA; 41469 MW; 9FC2F44539B8CC53 CRC64;
Query Match 5.1%; Score 105; DB 1; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.049;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VTVLLLAQTTCLLLFI-----ISRPGFPSPAGGEDRVH-----VLV 46
Db 31 IYIFFCFTICLLIFSSIKCKKLEQKLEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90
QY 47 LSSWRSGSFLGQLFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94
Db 91 GGVPMSGTTLMRAILDAAHPDVRCGGETMLLPFLTWQAGWRNDWV---NNSGIT----- 141
QY 95 DLMRSIFLCDMDVFDAYMPOSRNLSAFFNWATSRALCSPACSAFPRGTISK-QDVCKT 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167
QY 154 CTRQPFSLAREACRSYSHVVLKEVFFNLQVLYPLLSDPALNLRIVHLVRDPRAVLSR- 212
Db 168 CNKDP-----YTALWLPTIR-----RLYP-----NAKFILMIRARAVVHSMI 205
QY 213 EAAGPILARDNGIVLGTNGKWVEADPHLRILIREVCRSHVRIAEAAATLKPPFRLGRY 272
Db 206 ERKVPVAGYNTSDEISMEFQW---NQELRKMTFQCNN---APGQCIK-----V 247
QY 273 RFEDLAREPLAEIRALYAFTGLTLTPQLEAWTHNTHGSGIGKPI---EAPH TSS-RNA 327
Db 248 YYERLIQKPAEEILRITNFDLPFSQQM-----LRHQDLIGDEVLDNDQEFSA 301
QY 328 RNVS--QAWRHALPFTKILRVQEVCGALQLLGY-----RPVYS 364
Db 302 INTKALTSWFDCFSFEELRKLDLV-AFFIGILGYDTSISKPDYS 344

RESULT 3
YM67_MYCTU STANDARD; PRT; 388 AA.
AC Q50695;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 46.1 kDa protein Rv2267c.
GN Rv2267C OR MT2329 OR MTCY339.43.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; Pubmed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RT Nature 393:537-544(1998).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
```



```
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z77163; CAB00968.1; -.
DR EMBL; AE007076; AAK46611.1; -.
DR TIGR; MT2329; -.
DR TubercuList; Rv2267c; -.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 388 AA; 46062 MW; 5DED0263275A9B24 CRC64;

Query Match 5.0%; Score 102.5; DB 1; Length 388;
Best Local Similarity 22.7%; Pred. No. 0.091;
Matches 66; Conservative 37; Mismatches 99; Indels 89; Gaps 15;

QY 44 VLVLSSWRSGSFLGQLF---SQH-----PDVFYLMPEPAHWVWTTLSQGSAAATLH 90
Db 83 IFIVGHWTGTTLLHELLVVDRTGPTGYECLAPHHFLTE-----WFAPYVEFLVSKH 137
QY 91 MAVRDLMRSLFCLCDMDVFDAYMPQSRNLSAFFNWATSRALCSPPACSAFPRTGTSKQDVC 150
Db 138 RAMDNMDLSL-----HHPQDE----FVWC-MQGLSPYLTIAFP----- 172
QY 151 KTLCTRQP-----FSLAREACRSYSHVVLKEVRFNQLVLY-----PLLSDPALNLR I- 198
Db 173 -----NRPPQYEEYLDLEQVAPRELE--IWKRTLFRFVQVYFRRRKTVILKNPTHSFR I K 226
QY 199 -----VHLVRDPRAVLRREAAAGPILARDNGIVL-----GTNGKWEADPHRLRI 243
Db 227 VLEVEFPQAKFIHIVRDPVYVYPSTIHLKALYRIHGLQOPTFDGLDKVYST--YVDLY 284
QY 244 REVCRSHVRIAEATLKPPFLRGYRLVRFEELAREPLAEIRALYAFTGL 294
Db 285 R-----KLDEGRELVDP-----RFYELRYEDLLIGDEPQQLRLYQHLGL 324

RESULT 4
ODPT_HUMAN
ID ODPT_HUMAN STANDARD; PRT; 388 AA.
AC P29803;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pyruvate dehydrogenase E1 component alpha subunit, testis-specific
DE form, mitochondrial precursor (EC 1.2.4.1) (PDHE1-A type II).
GN PDHA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91065637; PubMed=2249846;
RA Dahl H.H.M., Brown R.M., Hutchison W.M., Maragos C., Brown G.K.;
RT "A testis-specific form of the human pyruvate dehydrogenase E1 alpha
RL subunit is coded for by an intronless gene on chromosome 4.";
RL Genomics 8:225-232(1990).
CC -!- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
CC CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
CC COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
CC DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
CC (E3).
CC -!- CATALYTIC ACTIVITY: Pyruvate + lipoamide = S-
```

```
acetyldihydrolipoamide + CO(2).
-!- COFACTOR: THIAMINE PYROPHOSPHATE.
-!- ENZYME REGULATION: E1 ACTIVITY IS REGULATED BY PHOSPHORYLATION
CC (INACTIVATION) AND DEPHOSPHORYLATION (ACTIVATION) OF THE ALPHA
CC SUBUNIT.
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- TISSUE SPECIFICITY: TESTIS. EXPRESSED IN POSTMEIOTIC SPERMATOGENIC
CC CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86808; AAA60232.1; -.
DR PIR; A37104; DEHUP1.
DR MIM; 179061; -.
DR InterPro; IPR001017; E1_dh.
DR Pfam; PF00676; E1_dehydrog; 1.
KW Glycolysis; Oxidoreductase; Flavoprotein; Thiamine pyrophosphate;
KW Phosphorylation; Mitochondrion; Transit peptide; Multigene family.
FT TRANSIT 1 27 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 28 388 PYRUVATE DEHYDROGENASE E1 COMPONENT
FT MOD_RES 230 230 ALPHA SUBUNIT, TESTIS-SPECIFIC FORM.
FT MOD_RES 291 291 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 298 298 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 388 AA; 42933 MW; 075B6CFF6DC73CC5 CRC64;

Query Match 4.7%; Score 96; DB 1; Length 388;
Best Local Similarity 23.2%; Pred. No. 0.35;
Matches 45; Conservative 25; Mismatches 82; Indels 42; Gaps 7;

QY 82 SQGS-AATLHMAVRDLMRSLFCLCDMDVFDAYMPQSRNLSAFFNWATSRALCSPPACSAFP 140
Db 198 NQGIATAEAFNMAALWKLPVFCICENNYG-----MGTSTERAAASP---DYK 242
QY 141 RGTI---SKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFNQLVLYPLLSDPALNLR 197
Db 243 RGNFIPGLKVDGMDVLCVREATKFAANYCRSGKGPILMELQYRYH--GHMSMDPGVSYR 300
QY 198 I-----VHLVRDPRAVLRREAAAGPILARDNGIVLGTNGKWEADPHRLRLIREVCRSHV 251
Db 301 TREETQEVRSKRDPITIIILQDRM-----VNSKLTAVEELKEIGAEVVRKEID 345
QY 252 RIAEAAATLKPPPF 265
Db 346 DAAQFATTDPEPHL 359

RESULT 5
PALL_IPOBA
ID PALL_IPOBA STANDARD; PRT; 707 AA.
AC P14166;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PAL.
OS Ipomoea batatas (Sweet potato) (Batate).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=4120;
RN [1]
RP SEQUENCE FROM N.A.
RA Tanaka Y., Matsuo M., Yamanoto N., Ohashi Y., Kano-Murakami Y.,
RA Ozeki Y.;
```



```
FT TRANSMEM 7 27 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 28 380 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 66 66 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 380 AA; 43313 MW; FF709BF00F1EDC95 CRC64;

Query Match 4.4%; Score 90.5; DB 1; Length 380;
Best Local Similarity 19.3%; Pred. No. 1.1;
Matches 66; Conservative 43; Mismatches 108; Indels 125; Gaps 17;

QY 51 RSGSFLGQLFSQHPDV-----FYLMEPAW----HVWTTLSQG--SAATLHMV 93
|||::: : |||: | : | : | : | : | : | : | : | : | : | : | : | :
79 RSGTILMRAMLDAHPEVRCGEETRVIPRIILNRSQWKSEKWNRLQQAGVTGEVINNAI 138
94 RDLMRSLFLCDMDVFDAYMPQSRNLSAFFNWTATSRALCSPACSAFPRGTISKQDVCKTL 153
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
139 SSFIMEIM-----VGHGDRAPRL 156
QY 154 CTRQPFSLAREACRSYSHVVVLKEVRFNQLVYLLSDPALNLRIVHLVRDPRAVLSRE 213
| : |||: | | ||| | : : : | : | : | : | : | : | : | : | : | : |
157 CNKDPPTMK-----SAVYLKE-----LFP-----NAKYLLMIRDGRATVNS-- 192
214 AAGPILARD--NGIVLG---TNGKWVEADPHLRLIREVCRSHVRIAEEAATLKPPPFELR 266
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
193 ----IISKVITIGFDLNDFRQCMTKWNAA---IQIMVDQCES---VGEKNCLK----- 236
267 GRYRLVREEDLAREPLAEIRALYAFTGLTLPQLEAWIHNIHNGSGIGKPIEAFHTSSRN 326
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
237 -----VYFEQLVLHPEAQMRRITEFLDI---PWDDKVLH---HEQLIGKDISLNVERSS 285
QY 327 ARNVSA-----WRHALPFTKILRVQEVCAQALQLLGYRP 361
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 286 DQVVKPVNLDALIKWGTIPEDVVADMDSV-APMLRRLGYDP 326

RESULT 9
CYB_LATCO
ID CYB_LATCO STANDARD; PRT; 371 AA.
AC Q9MLJ4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome B.
GN MTCYB OR COB OR CYTB.
OS Laticauda colubrina (Yellow-lipped sea krait).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20229584; PubMed=10764543;
RA Slowinski J.B., Keogh J.S.;
RT "Phylogenetic relationships of elapid snakes based on cytochrome b
mtDNA sequences."
RL Mol. Phylogenet. Evol. 15:157-164(2000).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF217834; AAF37253.1; -.
DR InterPro; IPR000179; Cyt_b_b6.
DR Pfam; PF00032; cytochrome_b_c; 1.
DR Pfam; PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00193; CYTOCHROME_B_OO; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; FALSE_NEG.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Heme.
FT METAL 75 75 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 89 89 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 174 174 IRON 2 (HEME B562 AXIAL LIGAND).
FT METAL 188 188 IRON 1 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 371 AA; 41947 MW; 36193FF6888D653C CRC64;

Query Match 4.3%; Score 89; DB 1; Length 371;
Best Local Similarity 23.3%; Pred. No. 1.4;
Matches 53; Conservative 32; Mismatches 76; Indels 66; Gaps 13;

QY 43 HVLVLSS-----WRSGSFLGQLFSQHPDVFYLMPEPAWHVWTTLSQGSAA TLHM 91
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 5 HTLLISNLLPVGSNISTWNFGSMLLSCLFLQTGTGFFL---AIHYTANINLAFSSVIHI 61
92 AVRDL-----MRSIFLCDMDVFDAYMPQSRNL--SAFFN---WATSRALCSP 133
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 62 -TRDVPYGWIMQNTHAISASAFFIC-----IYIHTARGLYGSLNKGWLTGVALLTT 114
134 PACSAF-----PRGTIS--KQDVCKTLC TRQPF---SLAREACRSYSHVVVLKEVRFENIQ 183
: || | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 115 LMATAFFGYVLPWGQMSFWAATVITNLLTAIPYLGTSLTTLWLGGSINDPTLTREFFALH 174
184 VLYPLLSDPALNLRIVHLVRDPAVLRSREAGPILARDNGIVLGTN 230
: | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 175 FILPFL---IISLSSTHII-----MLHNEGSSNP-----LGTN 204

RESULT 10
NODH_RHITR
ID NODH_RHITR STANDARD; PRT; 249 AA.
AC P52994;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Nodulation protein H (EC 2.8.2.-).
GN NODH.
OS Rhizobium tropici.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=398;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CFN 299;
RX MEDLINE=96303535; PubMed=8755625;
RA Laeremans T., Caluwaerts I., Verreth C., Rogel M.A.,
RA Vanderleyden J., Martinez-Romero E.;
RT "Isolation and characterization of Rhizobium tropici Nod factor
sulfation genes."
RL Mol. Plant Microbe Interact. 9:492-500(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CIAT899;
RX MEDLINE=97002746; PubMed=8850086;
RA Folch-Mallol J.L., Marroqui S., Sousa C., Manyani H., Lopez-Lara I.M.,
RA van der Drift K.M.G.M., Haverkamp J., Quinto C., Gil-Serrano A.,
RA Thomas-Oates J., Spaink H.P., Megias M.;
RT "Characterization of Rhizobium tropici CIAT899 nodulation factors: the
role of nodH and nodPQ genes in their sulfation."
RL Mol. Plant Microbe Interact. 9:151-163(1996).
CC -!- FUNCTION: REQUIRED FOR THE FORMATION OF SULFATED NOD FACTOR.
CC PROPOSED TO TRANSFER ACTIVATED SULFATE (PAPS) TO A N-
```


QY 208 V--LRSREAAGPILAR-----DNGIVLGTNGKWEADP---HLRLREVCRSHV 251
Db 1399 YERLLEMDLGCAGVSSKVDIPHSLAGRTHRGFAVGP-----DAGPGVIRLDRESVCY AHP 1453
QY 252 RIAEAATLKPPPELRGRYRLVRFFEDLAREPIAE-----IRALYAF----- 291
Db 1454 CLEE---LEFNAVYDS--ELVDISDMCCLPLATPKALFRPIYRSLSQSPRLALMDNYSFV 1508
QY 292 -----TGLTLTPQLEAWIHNITHGSGI-----GKPIEAFHTSSRNAR 328
Db 1509 MDLIMIRGLDIRPHLEEFDELLVVGQHILGQPVLVVEVYVGVVVRKRPVLAHRHPWSADLK 1568
QY 329 NVSQAWRHLPFTKILRVQEVCGALQLLG 358
Db 1569 RITVGGRAPCPSAARLR-DEDCQSSL-LVG 1596
RESULT 14
YPRB_CORGL
ID YPRB_CORGL STANDARD; PRT; 221 AA.
AC P46584;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in prob 5/region (Fragment).
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17965 / Melassecola;
RX MEDLINE=96345604; PubMed=8755867;
RA Ankri S., Serebrijski I., Reyes O., Leblon G.;
RT "Mutations in the Corynebacterium glutamicum proline biosynthetic
pathway: a natural bypass of the proA step.";
RL J. Bacteriol. 178:4412-4419(1996).
CC -!- SIMILARITY: SOME, TO B.SUBTILIS OBG C-TERMINAL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U31230; AAC44173.1; --
DR InterPro; IPR000765; GTP1_OBG.
DR PROSITE; PS00905; GTP1_OBG; PARTIAL.
KW Hypothetical protein.
FT NON_TER 1 1
SQ SEQUENCE 221 AA; 24716 MW; A0CA77DB6DF3DB02 CRC64;
Query Match 4.2%; Score 85.5; DB 1; Length 221;
Best Local Similarity 22.8%; Pred. No. 1.5;
Matches 53; Conservative 40; Mismatches 88; Indels 51; Gaps 12;
QY 162 AREACRSYSHVVLKEVRFNQLQVLYPLL-----SDPALNLRIVHLVRDPRAVLRSR 212
Db 14 APEALKSFAEVL--KVLFEKQFGWPVFIISAVARKALDP-LKYKLLLEIVQDARK-KRPK 69
QY 213 EAAGPILARDNGIVLGTNGKW-VEADPHLR---LI-----REVCRSHVRIAEAAATLKPP 262
Db 70 EKAESVILKPKAVVHTKQGQFIKPDPEVQGQFITGKPERWILQTFENDEAV----- 124
QY 263 PFLGRYRLVRFFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHG-----SGIGKPI 317
Db 125 GYLADRLSKLGIEDGLRKAGAHVGANVTIGGISF-----EWEPMTTAGDDPILTGRGTDV 179

QY 318 EAFHTSSRNARNVSVQAWRHLPFTKILRVQEVCGALQLLGYPVVSADQQR 369
Db 180 RLEQTS-----RISAAERK-----RASQVRRGLIDELDYGEDQEASRR 218
RESULT 15
ODPA_PIG
ID ODPA_PIG STANDARD; PRT; 389 AA.
AC P29804;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pyruvate dehydrogenase E1 component alpha subunit, somatic form,
DE mitochondrial precursor (EC 1.2.4.1) (PDH-E1-A type I) (Fragment).
GN PDH-E1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=90370488; PubMed=2395657;
RA Sermon K., Demelrleir L., Elpers I., Lissens W., Liebaers I.;
RT "Characterisation of a cDNA for porcine PDH-E1 alpha and comparison
RT with the human cDNA";
RL Nucleic Acids Res. 18:4925-4925(1990).
CC -!- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
CC CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
CC COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
CC DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
CC (E3).
CC -!- CATALYTIC ACTIVITY: Pyruvate + lipamide = S-
CC acetyldihydrolipoamide + CO(2).
CC -!- COFACTOR: THIAMINE PYROPHOSPHATE.
CC -!- ENZYME REGULATION: E1 ACTIVITY IS REGULATED BY PHOSPHORYLATION
CC (INACTIVATION) AND DEPHOSPHORYLATION (ACTIVATION) OF THE ALPHA
CC SUBUNIT.
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52990; CAA37180.1; --
DR PIR; S20813; DEPGPA.
DR InterPro; IPR001017; E1_dh.
DR Pfam; PF00676; E1_dehydrog; 1.
KW Glycolysis; Oxidoreductase; Flavoprotein; Thiamine pyrophosphate;
KW Phosphorylation; Mitochondrion; Transit peptide.
FT NON_TER 1 1
FT TRANSIT <1 28
FT CHAIN 29 389
FT
FT MOD_RES 231 231 MITOCHONDRION (BY SIMILARITY).
FT MOD_RES 292 292 PYRUVATE DEHYDROGENASE E1 COMPONENT
FT MOD_RES 299 299 ALPHA SUBUNIT, SOMATIC FORM.
FT MOD_RES 299 299 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 299 299 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 389 AA; 43121 MW; E9C7DF85369A9A47 CRC64;
Query Match 4.2%; Score 85.5; DB 1; Length 389;
Best Local Similarity 17.5%; Pred. No. 3.2;
Matches 47; Conservative 42; Mismatches 89; Indels 91; Gaps 7;
QY 38 GEDRVHVLVLSWSRSGSFLGQLESQH-----PDVFYLMPEAWHVWTTLSQGSAAATL 89
Db 184 GKDEV---CLTLYGDGAANQGQIFEAYNMAALWKLPCVFCENNRNRYMGTSVERAAASTD 240

QY	90	HMAVRDLMRSIFCLDCMDVFDAYMPQSRNLSAFFNWATSRALCSPPACSAFPRGTISKQDV	149
Db	241	YKRGDFIPGLRVDGMDI-----	258
QY	150	CKTLCTRQPFSLAREACRSYSHVVLKEVRFNQLVYLLSDPALNLRIVHLVRDPRAVL	209
Db	259	---LCVREATRFAAAYCRSGKPIMLQTYRYH-----GHMSMDPGVSY	300
QY	210	RSREAAGPILARDNGIVL---GTNGKWEADPHRLRIREVCRSHVRIAEATLKP----	261
Db	301	RTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFATADPEPPL	360
QY	262	-----PPF-LRGRYRLVRFEEDLA	278
Db	361	EELGYHIYCNDDPPFEVRGANQWIKFSIS	389

Search completed: September 20, 2002, 20:02:46
Job time: 255 sec

Db 336 RQPRWLGRYMLVRYEDVARAPLRKALEMYRFAGIHPTPQVEEWIRANTQAP---QDSNG 392

QY 320 FHTSSRNARNVSAQRHALPFTKILRVQEVGACALQLLGYRPVYSADQQRDLTLDLVLPR 379

Db 393 IYSTQKNSSEQPEKWRFSIPFKLAQVVQDACEPAMRLFYGYKLASSAQELNRSLSL-LEE 451

QY 380 GP 381

Db 452 GP 453

RESULT 5

E95934

probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sind

C;Species: Sinorhizobium meliloti

C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C;Accession: E95934

R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A;Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo

A;Reference number: A95842; MUID:21396508; PMID:11481431

A;Accession: E95934

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-307 <KUR>

A;Cross-references: GB:AL591985; PIDN:CAC49141.1; PID:gl5140626; GSPDB:GN00167

A;Experimental source: strain 1021, megaplasmid pSymb

R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A;Reference number: A96039; MUID:21368234; PMID:11474104

A;Contents: annotation

C;Genetics:

A;Gene: SMB21237

A;Genome: plasmid

RESULT 6

T16350

hypothetical protein F42G9.8 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000

C;Accession: T16350

R;Taich, A.

submitted to the EMBL Data Library, March 1996

A;Description: The sequence of C. elegans cosmid F42G9.

A;Reference number: Z18498

A;Accession: T16350

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-359 <TAI>

A;Cross-references: EMBL:U00051; NID:gl216305; PID:gl216308; PIDN:AAA91354.1; CESP:F4

A;Experimental source: strain Bristol N2

C;Genetics:

A;Gene: CESP:F42G9.8

A;Introns: 28/2; 50/2; 113/2; 155/3; 220/3; 290/2

C;Superfamily: Caenorhabditis elegans hypothetical protein F42G9.8

Query Match 5.1%; Score 105; DB 2; Length 359;

Best Local Similarity 19.8%; Pred. No. 0.14;

Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VTVLLLAQTTCLLLFI-----ISRPSPSPAGGEDRVH-----VLV 46

Db 31 IYIFIFCTICLLIFSSIKCKKLEQLESLSKESLIFNEQDARHSRLLSNLEQLIFV 90

QY 47 LSSWRSGSFELGQLFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94

Db 91 GGVPMSGTTLMRAILDADPDVRCGGFTMLPSFLTQAGWRNDWV---NNSGIT----- 141

QY 95 DLMRSIFLDMDFDAYMPQSRNLSAFFNWATSRALCSPACSAFPRGTISK-QDVCKTL 153

Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167

QY 154 CTRQPFSLAREACRSYSHVVLKEVREFNLQVLYPLLSDPALNLRIVHLVYRDPRAVLSR- 212

Db 168 CNKDP-----YTALWLPDIR-----RLYP-----NAKFILMTRDARAVVHSMI 205

QY 213 EAAGPILARDNGIVLGTNGKWEADPHLRILREVQSRSHVRIAEATLKPPFFLRGRYLV 272

Db 206 ERKVPVAGYNTSDEISMVFQW---NOELRKMFTQCNN---APGCQIK-----V 247

QY 273 RFEDLAREPLAEIRALYAFTGLTLTPQLBAWIHNITHGSGIGKPI-----EAFHTSS-RNA 327

Db 248 YYERLIQKPAEELLRIITNFDLDPFSQMQ-----LRHQDLIGDEVLDNQEFSSASQVKNS 301

QY 328 RNVS--QAWRHALPFTKILRVQEVGACALQLLGY-----RPVYS 364

Db 302 INTKALTSWFDCFSEETLRKLDV-APFLGILGYDTSISKPDYS 344

RESULT 7

G70729

hypothetical protein Rv2267c - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C;Accession: G70729

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A;Reference number: A70500; MUID:98295987

A;Accession: G70729

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-388 <COL>

Db 336 RQPRWLGRYMLVRYEDVARAPLRKALEMYRFAGIHPTPQVEEWIRANTQAP---QDSNG 392

QY 320 FHTSSRNARNVSAQRHALPFTKILRVQEVGACALQLLGYRPVYSADQQRDLTLDLVLPR 379

Db 393 IYSTQKNSSEQPEKWRFSIPFKLAQVVQDACEPAMRLFYGYKLASSAQELNRSLSL-LEE 451

QY 380 GP 381

Db 452 GP 453

RESULT 5

E95934

probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sind

C;Species: Sinorhizobium meliloti

C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C;Accession: E95934

R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A;Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo

A;Reference number: A95842; MUID:21396508; PMID:11481431

A;Accession: E95934

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-307 <KUR>

A;Cross-references: GB:AL591985; PIDN:CAC49141.1; PID:gl5140626; GSPDB:GN00167

A;Experimental source: strain 1021, megaplasmid pSymb

R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A;Reference number: A96039; MUID:21368234; PMID:11474104

A;Contents: annotation

C;Genetics:

A;Gene: SMB21237

A;Genome: plasmid

Query Match 5.2%; Score 107.5; DB 2; Length 307;

Best Local Similarity 20.6%; Pred. No. 0.067;

Matches 78; Conservative 46; Mismatches 131; Indels 123; Gaps 17;

QY 32 PSSPAGGEDRVHVLVSSW-RSGSSEFLGQLFSQHPDVF---YLMPEAWHVWTTLSQGSAA 87

Db 2 PSQP-----VRIAYIAGYGRSGSTILDIALGQHAADVVGAGEITSLTRHVW---RHNEYC 52

QY 88 TLHMAVRDLMRSIFLCD-----MDVFDAYMPQSRNLSAFFNWAISRAL 130

Db 53 ACNNAIRD-----CSFWSSVRREWSGDQDPLMEEYCALQKQFEGLSMMTRLSSGMGL 105

QY 131 CSPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFNQLQVLYPLLS 190

Db 106 -----GKQFSLYILHTKRLFS-AMQSC-SGRQVIVDSSKLPGRAMAVAQI- 148

QY 191 DPALNLRIVHLVRDPRAVLSRREAAGPILARDNGIVLGTNGKWEADPHLRILREVCRSH 250

Db 149 -PGIDMRVHILVRDGRGV-----AWSLL-----KGYERDAKSGLOKEI----- 185

QY 251 VRIAEAAATLKPPFFLRGRY-----LVRFEDLAREPLAEIRALY 289

Db 186 -----KPKSVFRTALRWSMVNLAVEYLSRKLKLGSEKVMRVRYEDFASDPVAVMQQIG 236

QY 290 APTGLTLTPQLEAWIHNITHGSGIGKPIEAFHTSSRN-----ARNVSAQRHALPFT 341

Db 237 TFELEDLS-QVGTSLN-----GEAMGPGHQVAGNRLRMNASIALNKDETWRTRMPAR 288

QY 342 KILRVQEVGACALQLLGY 359

Db 289 QQVSFQRLGGWMLRRYGY 306

Best Local Similarity 22.3%; Pred. No. 1.6;
Matches 69; Conservative 48; Mismatches 125; Indels 68; Gaps 17;
QY 33 SSPAGGEDRVHVLVSS-----WRSGSFLGQLFSQHDPDVFLMEPAWHVW 78
Db 70 SSNSGKLDTVSPMSSSDSDVFLSVALQYQGSGAALLSQFLKEHTRI--NPPYEGW 127
QY 79 T-TLSQGSAAATLHMAVRDL-----MRSIFL-----CDMDVFDAYMP 113
Db 128 NIIMTGTNTSCLDIALMLTNRGDSILVEKYSFPSALQSMRPLGLSCIPIDMDQF-GFLP 186
QY 114 QSRNLSAFPNW-ATSRALCSPPCSAFPRGTISKQDVCKTLCRQPFSLAREACRSYSHV 172
Db 187 ESM-DILTNWDATSYGSPKPHVLYTIPTGQNPSTGSTLSVERRKQIYTLAQK-----HDII 241
QY 173 VLKEVRFFNLQV-LY---PLLSDPALNLRIVHLVRDPRAVLRSREAAGPILARDN--GI 225
Db 242 ILEDEPYYYLOMDAYEGKPEADKAFTNE--QFVKELIPSLFSLMDVDGVRVIRMDLSKV 299
QY 226 VLGUNGKWEADPHRLRLIREVCRSHVRIAEAAATLKPPPFRLG-RYRLVRF--EDLAREPL 282
Db 300 AGRSRVGVWFTAQPLF-----IERGLRAAETATQSASGISQGIILYAMFKHWGQDGYLEWL 353
QY 283 AEIRALYAF 292
Db 354 KHIR--YSYT 361

RESULT 11
S78246
endopeptidase Clp (EC 3.4.21.-) ATP-binding chain clpC [similarity] - Odontella sinensis
N;Alternate names: ATP-dependent Clp proteinase regulatory chain; caseinolytic Clp prote
N;Contains: adenosinetriphosphatase (EC 3.6.1.3)
C;Species: Chloroplast Odontella sinensis
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 19-Jan-2001
C;Accession: S78246
R;Kowallik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.
Plant Mol. Biol. Rep. 13, 336-342, 1995
A;Title: The Chloroplast Genome of a chlorophyll a+c- containing Alga, Odontella sinensis
A;Reference number: S78238
A;Accession: S78246
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-885 <KOW>
A;Cross-references: EMBL:Z67753; NID:g1185127; PIDN:CAA91619.1; PID:g1185136
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995
C;Genetics:
A;Gene: clpC
A;Genome: chloroplast
C;Function:
A;Description: allows clpp to hydrolyze polypeptides and proteins, probably by a chaperon
e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller
C;Superfamily: endopeptidase Clp ATP-binding chain
C;Keywords: ATP; chloroplast; hydrolase; molecular chaperone; nucleotide binding; P-loop
F;218-225/Region: nucleotide-binding motif A (P-loop)
F;285-290/Region: nucleotide-binding motif B
F;560-567/Region: nucleotide-binding motif A (P-loop)
F;628-633/Region: nucleotide-binding motif B
F;224/Binding site: ATP (Lys) #status predicted
F;566/Binding site: ATP (Lys) #status predicted

Query Match 4.6%; Score 95; DB 1; Length 885;
Best Local Similarity 21.0%; Pred. No. 3.6;
Matches 89; Conservative 59; Mismatches 161; Indels 114; Gaps 18;
QY 5 RESSKTVTVLLAQ-----TTCLLLFII-SRPGSPSPAGGEDRVHVLVLSWR 51
Db 4 KFTGALKVIMLSQEEARMGMHNFVGTETQQLLIGIQRHGIGARALKQKV---TLKKAR 60
QY 52 -----SGSSFLGQLFSQHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLMSR--IF 101
Db 61 REIELYIGRTGTFVASEIPTFPRAKRVLMAVHEGKDLGQNFVGTETHEILLALISESDGVA 120

QY 102 LCDMDVFDAYMPQSRNLSAFFNWTATSRALCSPPCSAFPRGTISKQDVCKTLCRQPFSL 161
Db 121 MRTLDKLGVNIPKLRNLILMYI-----EENQEEILRPLTQAEKFL 161
QY 162 AREACRS-----YSHVVLKEVRFFNLQVLYPLSDPALNL-RIVHLVRDPRAVLRSRE 213
Db 162 EREKKGSTPTLDEYSENISKEAVDGKL-----DPVIGRDKEIHEV---IKVLARRR 210
QY 214 AAGPILARDNGIVLTNGKWEADPHRLRLIREVCRSHVRIAEAAATLKPPPFRLGRYLRV 273
Db 211 KNPVLIGEPGV-----GKTAVAEGLAQLI-----IAE-----KAPDFLDGNL---- 248
QY 274 FEDLAREPLAEIRALYAF 273
Db 249 ---LMALDLSILAGTKYRG-EFEERIKRIVEEVQND 304
QY 329 NVSQAWRHLPFTKILRVQEV 374
Db 305 DAANILKPALARGKFR-----CIGATTIDEVRKYIERDPALERRFQPVHKEPTVGV 359
QY 375 LVL 377
Db 360 ILL 362

RESULT 12
VGWJBV
peplomer glycoprotein precursor - Berne virus (strain p138/72)
C;Species: Berne virus
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999
C;Accession: A36759
R;Snijder, E.J.; Den Boon, J.A.; Spaan, W.J.M.; Weiss, M.; Horzinek, M.C.
Virology 178, 355-363, 1990
A;Title: Primary structure and post-translational processing of the Berne virus pepl
A;Reference number: A36759; MUID:91020973
A;Accession: A36759
A;Molecule type: genomic RNA
A;Residues: 1-1581 <SNI>
A;Cross-references: GB:X52506; NID:g62059; PIDN:CAA36748.1; PID:g62050
C;Genetics:
A;Gene: P
C;Superfamily: Berne virus peplomer glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1581/Product: peplomer glycoprotein #status predicted <PGP>
F;1547-1572/Domain: transmembrane #status predicted <TMN>
F;25,310,384,494,574,935,969,1267,1297,1327,1385,1389,1428,1431,1438,1483,1487,1495,1

Query Match 4.6%; Score 94.5; DB 1; Length 1581;
Best Local Similarity 17.6%; Pred. No. 8.7;
Matches 81; Conservative 65; Mismatches 135; Indels 179; Gaps 19;
QY 2 WLPRFSSKTVTVLLLAQTCLLLFIISRPGSPSPAGGEDRVHVLVLSWRSGSFLGQLF 61
Db 259 WVVAFQNKATAVILPSELIVPVAQKVTR-----RLG 289
QY 62 SQHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLMSRIFLCMDVFDAYMPQSRNLSAF 121
Db 290 VNTPDYFWLVKQAYH-----YLSQA-NLSP- 313
QY 122 FNWATSRALCSPPCSAFPRGTISKQDVCKTLCRQPFSLAREACRSYSHV-LKEVREF 180
Db 314 -NYALFSALCN-----SLYQQSATLSTLCFSGPFFVAQEC---YNNALYLPDAVFT 360
QY 181 NL-----QVLYPL-----LSDPALNL-----RIVHLVRDPRAVL---- 209
Db 361 TLFTSLFSWDYQINYPNLNQVLTQNETFLQLPATNYQGQTLQSGRMLNLFKDAIVFLDFFD 420
QY 210 -----RSREAAGP-----ILARDNGIVLGTNGKWEADPHRLRLIREVCRSHVRIAEAAATLK 261
Db 421 TKFYENDAPSSDIFVWVARQAQLIRYGNFRIEQINGYFQV---KCSSNI-----ISTLEP 473

QY 262 PPFLRGYRLVRFEDLAREPLAEIRALYAFT----- 292

Db 474 HP--AGVIMIAHHHSMWSVAARNSTSFYCVTHSLTTFGKLDISTSWFFHTLALSPVSQ 531

QY 293 -----GLILTPQLEAWIHNIT-----HGSGIGKPIEAFHTSSRNARNVS 331

Db 532 VSMPLLLSTAAGVVMHPMIEHWIPLLTLAQSQYQPSFFNIGINKTI-TLTTQLQAYAQVY 590

QY 332 QAWRHALPFTKILRVQEVCAQALQLLGYRVPVYSADQQRDIL 371

Db 591 TAWFLSVIYVRLPEARRLILG-VQLVPFIQALLSIKQADL 629

RESULT 13

A70841

probable oxalyl-CoA decarboxylase - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C;Accession: A70841

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987

A;Accession: A70841

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-582 <COL>

A;Cross-references: GB:AL021926; GB:AL123456; NID:g3261520; PIDN:CAA17312.1; PID:g290958

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: oxCA

C;Superfamily: oxalyl-CoA decarboxylase; thiamin pyrophosphate-binding domain homology

Query Match 4.6%; Score 94; DB 2; Length 582;

Best Local Similarity 21.4%; Pred. No. 2.6;

Matches 55; Conservative 44; Mismatches 98; Indels 60; Gaps 11;

QY 152 TLCTROPFSLAREACRSYSHVVLKEVRFENLQVLYPLLSDP-----ALNLRIVHLV 202

Db 2 TTRSASPTVLTDC-----HLVVDALKANDVDITYGVVGIPITDLARAQAQSGIRYIGF- 56

QY 203 RDPRAVLRSREAAGPILARDNGIVLGTNGK-WVEADPHRLRIREVCRSHVRIAEAAATLKP 261

Db 57 RHEASAGNAAAAGFLTARP-GVCLTTSGPGLNLPALANATNCFPMIQISGSSS-RP 114

QY 262 -----PPFLRGYRLVRFEDLAREPLAEIRALYAFTGLILTPQLEAW 303

Db 115 MVDLQRGDYQDLQDNAARPFVKAAYRIGOVQDIDRGVARAIR-----TATSGRPGG 166

QY 304 IHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCAQALQLLGYRVPY 363

Db 167 VYLDIPGDVLGQAVEASAAS-----GAIWRPVDPAPELLPAPEAIDRALDVL----- 213

QY 364 SADQQRDLTLDVLPRG 380

Db 214 -AQAQRPL---LVLSKG 226

RESULT 14

C75278

DNA polymerase III, tau/gamma subunit - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C;Accession: C75278

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896

A;Accession: C75278

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-615 <WHI>

A;Cross-references: GB:AE002071; GB:AE000513; NID:g6460218; PIDN:AAF11953.1; PID:g646

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR2410

A;Map position: 1

C;Superfamily: DNA-directed DNA polymerase III gamma chain

Query Match 4.6%; Score 93.5; DB 2; Length 615;

Best Local Similarity 23.4%; Pred. No. 3.1;

Matches 92; Conservative 41; Mismatches 141; Indels 119; Gaps 21;

QY 4 PRFSSKTYTVILLAAQTCLLLFIISRPGPSSPAGGEDRVHVLVLSWSRSGSFLGQLFSQ 63

Db 44 PRGVGKTTTARLIAMTA-----NCTGPAPKPCGECECLAV-----RAGS----- 83

QY 64 HPDVFYLMERPAWHVWTTLSQGSAAATLHMA--VRDLMRISIFLCDMD-----VFDAYMPQS 115

Db 84 HPDVMEI-----DAASNNSVDVDRDLREKVGLAAMRGCKKIYILDEAHMMS 129

QY 116 RNLSAFFNWAISRALCSPSPACSAFFPRGTISKQDVCKTLCRQPFSLAREACRSY--SHVV 173

Db 130 R---AAFN-ALLKTLLEPPPEHVIFILATTEPEKIPTILSR-----CQHYRFRRLT 176

QY 174 LKEV--RFFNLQVLYPLLSDP-ALNL--RIVH-LVRDPRAVLRSREAAGPILARDNGIVL 227

Db 177 SEEIAGKLAGLVLEGASADPDALNLIGRLADGAMRDGESLLERMLAAGTAVTRP----- 231

QY 228 GTNGKWEADPHRLRIREVCRSHVRIAEATLKPPPLRGYRLVRFEDLAREPLAEIRA 287

Db 232 -----AVEEALGPLPGERVGVASALLVGD-AGEAISGAAQ 266

QY 288 LY--AFTGLTLTPQLEAWIHNITHGS-GIGKPIEAFHTSSRNARNVSQAWRHALPFTKIL 344

Db 267 LYRDGFAARTVVEGLVAAFGAALHAELGLGE-----EGRLEGAEVPRLL 310

QY 345 RVQEVCAQALQLLGYRVPVYSADQQRDLTLDLVL 377

Db 311 KLQ----AALDEQEAREFARSADQQ---SLELAL 336

RESULT 15

B83629

probable ATP-binding component of ABC transporter PA0136 [imported] - Pseudomonas aeruginosa

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C;Accession: B83629

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V. Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

A;Reference number: A82950; MUID:20437337

A;Accession: B83629

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-523 <STO>

A;Cross-references: GB:AE004451; GB:AE004091; NID:g9945958; PIDN:AAG03526.1; GSPDB:G

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA0136

Query Match 4.5%; Score 93; DB 2; Length 523;

Best Local Similarity 26.4%; Pred. No. 2.7;

Matches 37; Conservative 17; Mismatches 60; Indels 26; Gaps 5;

QY 13 VLLLAQTTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQLFSQHDPDVFYLM 72
Db 384 VLALAE-EIIRREAVKAPGAGAPARS-----LGGNLQKFILGREILQAPRLVAAH 434
QY 73 PAWHVWTTLSQSSAATLH---MAVRDLMRSIFLCDDMDVFDAYMPQSRNLSAFENWATSRA 129
Db 435 PTWGV----DVGAALIHREALRDAGTALLVVSEDLDELFLLSDR-----IAA 480
QY 130 LCSPPACSAFPRGTISKQDV 149
Db 481 LCSGRCLCPAVATASASPOKV 500

Search completed: September 20, 2002, 19:58:58
Job time: 147 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 19:56:06 ; Search time 14.8 Seconds
(without alignments)
643.647 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRFSSKTVTVLLLAQTT.....LTLDLVLPGRGPDHFSWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1008	49.1	386	4	US-09-045-284A-2
2	665	32.4	483	3	US-09-263-023-2
3	656.5	32.0	484	3	US-09-263-023-4
4	577.5	28.2	479	2	US-08-899-514-2
5	540.5	26.4	458	2	US-08-655-878-2
6	105	5.1	359	3	US-09-150-133-11
7	105	5.1	359	3	US-09-150-141-11
8	105	5.1	359	4	US-09-374-493-11
9	105	5.1	359	4	US-09-374-824-11
10	105	5.1	359	4	US-09-374-492-11
11	90.5	4.4	380	3	US-09-150-133-9
12	90.5	4.4	380	3	US-09-150-141-9
13	90.5	4.4	380	4	US-09-374-493-9
14	90.5	4.4	380	4	US-09-374-824-9
15	90.5	4.4	380	4	US-09-374-492-9
16	90.5	4.4	984	2	US-08-673-789-9
17	86.5	4.2	376	3	US-09-150-133-7
18	86.5	4.2	376	3	US-09-150-141-7
19	86.5	4.2	376	4	US-09-374-493-7
20	86.5	4.2	376	4	US-09-374-824-7
21	86.5	4.2	376	4	US-09-374-492-7
22	86.5	4.2	377	3	US-09-150-133-5
23	86.5	4.2	377	3	US-09-150-141-5
24	86.5	4.2	377	4	US-09-374-493-5
25	86.5	4.2	377	4	US-09-374-824-5
26	86.5	4.2	377	4	US-09-374-492-5
27	86.5	4.2	566	2	US-08-484-993B-41

28	86.5	4.2	566	2	US-08-484-158B-41	Sequence 41, Appl
29	86.5	4.2	566	2	US-08-484-596A-41	Sequence 41, Appl
30	86.5	4.2	566	2	US-08-480-150A-41	Sequence 41, Appl
31	86.5	4.2	566	3	US-08-458-731-41	Sequence 41, Appl
32	86.5	4.2	566	3	US-08-149-223A-41	Sequence 41, Appl
33	86.5	4.2	1711	3	US-08-369-822C-10	Sequence 10, Appl
34	86.5	4.2	1711	3	US-08-582-776C-10	Sequence 10, Appl
35	86.5	4.2	1711	3	US-08-434-831B-10	Sequence 10, Appl
36	83.5	4.1	390	4	US-09-108-020-36	Sequence 36, Appl
37	83	4.0	15281	2	US-08-471-119A-2	Sequence 2, Appl
38	81.5	4.0	829	4	US-08-444-818-69	Sequence 69, Appl
39	81.5	4.0	1786	4	US-08-444-818-54	Sequence 54, Appl
40	81.5	4.0	2261	4	US-08-444-818-66	Sequence 66, Appl
41	81.5	4.0	2436	4	US-08-444-818-75	Sequence 75, Appl
42	81.5	4.0	2772	4	US-08-444-818-89	Sequence 89, Appl
43	81.5	4.0	2894	2	US-08-466-975A-23	Sequence 23, Appl
44	81.5	4.0	2894	2	US-08-391-671A-23	Sequence 23, Appl
45	81.5	4.0	2894	3	US-08-467-902A-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match 49.1%; Score 1008; DB 4; Length 386;
Best Local Similarity 52.3%; Pred. No. 1.1e-102;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY	14	LLLAQTTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSSWSRSGSFLGQ	59
Db	1	MLLPKKMKLLFLVSQMAILALFFHMYSHNSSLKAKQPERMHVVLVLSSWSRSGSFLGQ	60
QY	60	LFSQHPDVFYLMEPANWHVWTTLSQGSAA TLHMAVRDLMSIFLCMDVFDAYM-PQSRNL	118
Db	61	LFGQHPDVFYLMEPANWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRRQ	120
QY	119	SAFENWATSRALCSPACSAFPRGTISKQDVCKILCTRPFSRLAREACRSYSHVVLKEVR	178
Db	121	SSLFQWENSRALCSAPACDIIPQDEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVR	180
QY	179	FFNLQVLYPLLSDFALNLRIVHLVRDPRAVLSREAAGPILARDNGIVLGTN-GKWVEAD	237
Db	181	FFNLQSLYPLLKDPSSLNHLVHLVRDPRAVLSREAAGPILARDNGIVLGTN-GKWVEAD	240
QY	238	PHLRLIREVCRSHVRIAEEAATLKP-PPFLRGYRLVRFEFLAREPLAEIRALYAFTGLTL	296
Db	241	OPYVVMQVICSQLEIYK--TIQSLPKALQRYLLVRYEDLARAPVACTSRMYEFGLEF	298
QY	297	TPQLEAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCAQALQL	356
Db	299	LPHLQTWVHNIRGKMGMD--HAFHTNARDALNVSQAWRWSLPYEKVSRLOKACGDAMNL	356
QY	357	LGYRPVYSADQQQDRLTLDLVLPGRGPDHFSWASPD	390

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,514
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DANIEL E ALTMAN
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714 760 0404
TELEFAX: 714 760 9502
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 479
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-899-514-2

Query Match 28.2%; Score 577.5; DB 2; Length 479;
Best Local Similarity 35.7%; Pred. No. 4.7e-55;
Matches 132; Conservative 65; Mismatches 134; Indels 39; Gaps 8;

QY 32 PSSPAGGEDRVHVLVLSWRSGLQFSGHDPDVFYLMPEAWHVWTTLS---QGSAA 87
Db 123 PPRPAVAGPRRHVLLMATRTGSSVGEFFNQGNIFYLFEPLWHIERTVSFEPGGANAA 182
QY 88 TLHMAVRDLMSIFLCMDVFDAY---MPQSRNLSAFFNWTATSRALCSPACSAFPRGTI 144
Db 183 GSALYRDVLKQLFLCDLVLEHFIITPLPEDHLTQEMFRGSSRLCEDPVCTPFVKVF 242
QY 145 SKQDVCKTLCTRPFSLAAREACRSYSHVVLKEVRFNQLVLYPLSDPALNLRIVHLVRD 204
Db 243 EKYHCKNRRCGPNLTLAAEACRRKEHMAKAVRIRQLFEFLQPLAEDPRDLRLVQLVRD 302
QY 205 PRAVLRSREAAGPILARDNGIVLG---TNGKWE-----ADPHLRLIREVCRSHVRIA 254
Db 303 PRAVLASRMVA-----FAGKYKTWKKWLDDEGQDGLREEVQRLRGNCES-IRLS 351
QY 255 EAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHNITH---G 310
Db 352 AELGLRQPAWLGRYMLVRYEDVARGPLQKAREMYPFAGIPLTPQVEDWIKNTQAAHDG 411
QY 311 SGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYRPPVSADQQRD 370
Db 412 SGI-----YSTQKNSSEQEKEKWRFSMPFKLAQVVQAPCGPAMRLFGYKLARDAAALTN 464
QY 371 LTLDLVLPRG 380
Db 465 RSVSLLEERG 474

RESULT 5
US-08-655-878-2
Sequence 2, Application US/08655878
Patent No. 5827713
GENERAL INFORMATION:
APPLICANT: FUKUTA, MASAKAZU
APPLICANT: HABUCHI, OSAMI
TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE:
STREET:

CITY:
STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,878
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE:
TELEFAX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 458
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-655-878-2

Query Match 26.4%; Score 540.5; DB 2; Length 458;
Best Local Similarity 34.3%; Pred. No. 5.3e-51;
Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;

QY 36 AGGEDRVHVLVLSWRSGLQFSGHDPDVFYLMPEAWHV---WTTLSQGSAAATLHMAY 93
Db 107 AAPEPRRHVLLMATRTGSSVGEFFNQGNIFYLFEPLWHIERTVTFEPGGANAVGSAL 166
QY 94 --RDLMSIFLCMDVFDAYM---PQSRNLSAFFNWTATSRALCSPACSAFPRGTISKQD 148
Db 167 VYRDVLQQLLLCDLYILESFISPADEEHLTAALFRRGSSHSLCEEPVCTPSLKKVFEKYH 226
QY 149 VCKTLCTRPFSLAAREACRSYSHVVLKEVRFNQLVLYPLSDPALNLRIVHLVRDPRAV 208
Db 227 CKNRCCGPNLITLAAEACRRKQHMALKTVRIRQLFEFLQPLAEDPRDLRLIQLVRDPRAV 286
QY 209 LRSREAAGPILARDNGIVLGNGKWE-----ADPHLRLIREVCRSHVRIAEATL 259
Db 287 LVSRMVA-----FSGKYESWKKWAAEGEAPLOEDEVQRLRGNCES-IRLSAELGL 335
QY 260 KPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIHNITHGSGICKPIEA 319
Db 336 RQPRWLGRYMLVRYEDVARAPLRKALEMYRFAGIHFTPQVEEWIRANTQAP---QDSNG 392
QY 320 FHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYRPPVSADQQRDLTLDLVLP 379
Db 393 IYSTQKNSSEQEKEKWRFSIPFKLAQVVQDACEPAMRLFGYKLASSAQELTNRSLSL-LEE 451
QY 380 GP 381
Db 452 GP 453

RESULT 6
US-09-150-133-11
Sequence 11, Application US/09150133B
Patent No. 6060295
GENERAL INFORMATION:
APPLICANT: The Board of Regents of the University of Oklahoma
TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF

```

; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-133-11

Query Match          5.1%; Score 105; DB 3; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0041;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VTVLLLAQTTCLLLFI-----ISRPGSPSPAGGEDRVH-----VLV 46
Db 31 IYIFFCFTICLLIFSSIKCKKLEQKLEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90

QY 47 LSSWRSGSFLGQLFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94
Db 91 GGVPRSGTILMRAILDADHPDVRCGGETMLLPFLTWQAGWRNDWV---NNSGIT----- 141

QY 95 DLMRSIFLCDMDVFDAYMPPQSRNLSAFFNWATSRALCSPACSAFPRGTISK-QDVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167

QY 154 CTRQPFSLAREACRSYSHVVLKEVRFENLQVLYPLLLSDPALNLRIVHLVRDPRAVLRSR- 212
Db 168 CNKDP-----YTALWLPTIR-----RLYP-----NAKFILMIRDARAVVHSMI 205

QY 213 EAAGPILARDNGIVLGTNGKWEADPHLRILIREVCRSHVRIAEAAATLKPPPFRLGRYRLV 272
Db 206 ERKVPVAGYNTSDEISMVFQW---NQELRKMTFQCNN----APGQCIK-----V 247

QY 273 RFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHGSGIGKPI----EAFHTSS-RNA 327
Db 248 YYERLIQKPAEEILRITNFDLPFSQQM-----LRHQDLIGDEVLDNDQEFASQVKNS 301

QY 328 RNVS--QAWRHALPFTKILRVQEVCAQALQLLGY-----RPVYS 364
Db 302 INTKALTSWFDCFSSEETLRKLDV-APFLGILGYDTSISKPDYS 344

RESULT 7
US-09-150-141-11
; Sequence 11, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-141-11
```

```

Query Match          5.1%; Score 105; DB 3; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0041;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VTVLLLAQTTCLLLFI-----ISRPGSPSPAGGEDRVH-----VLV 46
Db 31 IYIFFCFTICLLIFSSIKCKKLEQKLEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90
```

```

QY 47 LSSWRSGSFLGQLFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94
Db 91 GGVPRSGTILMRAILDADHPDVRCGGETMLLPFLTWQAGWRNDWV---NNSGIT----- 141

QY 95 DLMRSIFLCDMDVFDAYMPPQSRNLSAFFNWATSRALCSPACSAFPRGTISK-QDVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167

QY 154 CTRQPFSLAREACRSYSHVVLKEVRFENLQVLYPLLLSDPALNLRIVHLVRDPRAVLRSR- 212
Db 168 CNKDP-----YTALWLPTIR-----RLYP-----NAKFILMIRDARAVVHSMI 205

QY 213 EAAGPILARDNGIVLGTNGKWEADPHLRILIREVCRSHVRIAEAAATLKPPPFRLGRYRLV 272
Db 206 ERKVPVAGYNTSDEISMVFQW---NQELRKMTFQCNN----APGQCIK-----V 247

QY 273 RFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHGSGIGKPI----EAFHTSS-RNA 327
Db 248 YYERLIQKPAEEILRITNFDLPFSQQM-----LRHQDLIGDEVLDNDQEFASQVKNS 301

QY 328 RNVS--QAWRHALPFTKILRVQEVCAQALQLLGY-----RPVYS 364
Db 302 INTKALTSWFDCFSSEETLRKLDV-APFLGILGYDTSISKPDYS 344

RESULT 8
US-09-374-493-11
; Sequence 11, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374,493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-493-11
```

```

Query Match          5.1%; Score 105; DB 4; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0041;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VTVLLLAQTTCLLLFI-----ISRPGSPSPAGGEDRVH-----VLV 46
Db 31 IYIFFCFTICLLIFSSIKCKKLEQKLEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90

QY 47 LSSWRSGSFLGQLFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94
Db 91 GGVPRSGTILMRAILDADHPDVRCGGETMLLPFLTWQAGWRNDWV---NNSGIT----- 141

QY 95 DLMRSIFLCDMDVFDAYMPPQSRNLSAFFNWATSRALCSPACSAFPRGTISK-QDVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167

QY 154 CTRQPFSLAREACRSYSHVVLKEVRFENLQVLYPLLLSDPALNLRIVHLVRDPRAVLRSR- 212
Db 168 CNKDP-----YTALWLPTIR-----RLYP-----NAKFILMIRDARAVVHSMI 205

QY 213 EAAGPILARDNGIVLGTNGKWEADPHLRILIREVCRSHVRIAEAAATLKPPPFRLGRYRLV 272
Db 206 ERKVPVAGYNTSDEISMVFQW---NQELRKMTFQCNN----APGQCIK-----V 247
```



```
QY 273 RFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHGSGIGKPI-----EAFHTSS-RNA 327
Db 248 YYERLIQKPAEEILRITNFDLPFSQQM-----LRHQDLIGDEVLDNDQEFASQVKNS 301
QY 328 RNVS--QAWRHALPFTKILRVQEVCGALQLLGY-----RPVYS 364
Db 302 INTKALTSWFDCFSSEETLRKLDVV-APFLGILGYDTSISKPDYS 344
```

RESULT 9

```
US-09-374-824-11
; Sequence 11, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; EARLIER FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-824-11
```

```
Query Match 5.1%; Score 105; DB 4; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0041;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;
```

```
QY 11 VTVLLLAQTTCLLLFI-----ISRPGPSSPAGGEDRVH-----VLV 46
Db 31 IYIFIFCFTICLLIFSSIKCKKLEQLSLSKESLIFNEQDARHSRRLLSNLEQLIFV 90
QY 47 LSSWRSGSFLGQLFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94
Db 91 GGVPRSGTTLMAILDAHPDVRCGGETMLLPSELTWQAGWRNDWV---NNSGIT----- 141
QY 95 DLMRSIFLCDMDVFDAYMPQSRNLSAFFNWATSRALCSPACSAFPRGTISK-QDVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167
QY 154 CTRQPFSLAREACRSYSHVVLKEVRFENLQVLYPLLSDPALNLRIVHLVRDPRAVLSR- 212
Db 168 CNKDP-----YTALWLPTIR-----RLYP-----NAKFILMIRDARAVVHSMI 205
QY 213 EAAGPILARDNGIVLTNGKWEADPHLRILIREVCRSHVRIAEATLKPPPFRLGRYRLV 272
Db 206 ERKVPVAGYNTSDELSMFVQW---NQELRKMTFQCNN---APGQCICK-----V 247
QY 273 RFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHGSGIGKPI-----EAFHTSS-RNA 327
Db 248 YYERLIQKPAEEILRITNFDLPFSQQM-----LRHQDLIGDEVLDNDQEFASQVKNS 301
QY 328 RNVS--QAWRHALPFTKILRVQEVCGALQLLGY-----RPVYS 364
Db 302 INTKALTSWFDCFSSEETLRKLDVV-APFLGILGYDTSISKPDYS 344
```

RESULT 10

```
US-09-374-492-11
; Sequence 11, Application US/09374492
; Patent No. 6207432
; GENERAL INFORMATION:
```

```
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.545
; CURRENT APPLICATION NUMBER: US/09/374,492
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,141
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-492-11
```

```
Query Match 5.1%; Score 105; DB 4; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0041;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;
```

```
QY 11 VTVLLLAQTTCLLLFI-----ISRPGPSSPAGGEDRVH-----VLV 46
Db 31 IYIFIFCFTICLLIFSSIKCKKLEQLSLSKESLIFNEQDARHSRRLLSNLEQLIFV 90
QY 47 LSSWRSGSFLGQLFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94
Db 91 GGVPRSGTTLMAILDAHPDVRCGGETMLLPSELTWQAGWRNDWV---NNSGIT----- 141
QY 95 DLMRSIFLCDMDVFDAYMPQSRNLSAFFNWATSRALCSPACSAFPRGTISK-QDVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167
QY 154 CTRQPFSLAREACRSYSHVVLKEVRFENLQVLYPLLSDPALNLRIVHLVRDPRAVLSR- 212
Db 168 CNKDP-----YTALWLPTIR-----RLYP-----NAKFILMIRDARAVVHSMI 205
QY 213 EAAGPILARDNGIVLTNGKWEADPHLRILIREVCRSHVRIAEATLKPPPFRLGRYRLV 272
Db 206 ERKVPVAGYNTSDELSMFVQW---NQELRKMTFQCNN---APGQCICK-----V 247
QY 273 RFEDLAREPLAEIRALYAFTGLTLTPQLEAWIHNITHGSGIGKPI-----EAFHTSS-RNA 327
Db 248 YYERLIQKPAEEILRITNFDLPFSQQM-----LRHQDLIGDEVLDNDQEFASQVKNS 301
QY 328 RNVS--QAWRHALPFTKILRVQEVCGALQLLGY-----RPVYS 364
Db 302 INTKALTSWFDCFSSEETLRKLDVV-APFLGILGYDTSISKPDYS 344
```

RESULT 11

```
US-09-150-133-9
; Sequence 9, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 9
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-133-9
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 19:55:26 ; Search time 32.41 Seconds
(without alignments)
1336.587 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRFSSKTVTVLLAQTT.....LTLDLVLPRGPDHFSWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2051	100.0	390	AA72639	Human glycosyl sul
2	1865.5	91.0	418	AA41947	Human OREF ORF1711
3	1729.5	84.3	395	AA72640	Human glycosyl sul
4	1536.5	74.9	395	AA72638	Mouse glycosyl sul
5	1013	49.4	386	AA93309	Human polypeptide,
6	1008	49.1	386	AA39918	Human glycosyl sul
7	970	47.3	386	AA79219	Human transferase
8	958.5	46.7	388	AA39919	Mouse glycosyl sul
9	665	32.4	483	AA31656	Mouse N-acetylgluc
10	659.5	32.2	530	AA95367	Human protein sequ
11	656.5	32.0	484	AA31657	Human N-acetylgluc

12	577.5	28.2	479	19	AAW52863	Glycosaminoglycan
13	540.5	26.4	458	18	AAW06480	Chick chondroitin
14	513.5	25.0	411	19	AAW61100	Keratan sulphate 6
15	184.5	9.0	183	22	ABB68582	Drosophila melanog
16	174.5	8.5	363	22	ABB64512	Drosophila melanog
17	119.5	5.8	596	22	AA72641	Human glycosyl sul
18	119.5	5.8	1222	22	AA72642	Human glycosyl sul
19	112.5	5.5	1207	22	AA72643	Mouse glycosyl sul
20	107.5	5.2	315	22	ABB64513	Drosophila melanog
21	105	5.1	359	20	AA706628	C. elegans tyrosyl
22	105	5.1	359	21	AA784309	A tyrosylprotein s
23	99.5	4.9	928	22	AAE05075	Drosophila melanog
24	99.5	4.9	972	22	ABB63015	Drosophila melanog
25	96	4.7	692	22	ABG15672	Novel human diagno
26	91.5	4.5	953	22	ABG03420	Novel human diagno
27	90.5	4.4	380	20	AA706627	C. elegans tyrosyl
28	90.5	4.4	380	21	AA784308	A tyrosylprotein s
29	89	4.3	332	22	ABG06686	Novel human diagno
30	88	4.3	268	22	AA48176	Thermus thermophil
31	87.5	4.3	664	22	ABB70395	Drosophila melanog
32	87	4.2	632	22	ABB63316	Drosophila melanog
33	86.5	4.2	330	21	AA34913	Gene 7 human secre
34	86.5	4.2	330	21	AA34914	Human secreted pro
35	86.5	4.2	376	20	AA706626	Mouse tyrosylprote
36	86.5	4.2	376	21	AA784307	A murine tyrosylpr
37	86.5	4.2	377	20	AA706625	Human tyrosylprote
38	86.5	4.2	377	21	AA784306	A human tyrosylpro
39	86.5	4.2	377	21	AA769421	Amino acid sequenc
40	86.5	4.2	377	22	AA93219	Human polypeptide,
41	86.5	4.2	377	22	AA93565	Human polypeptide,
42	86.5	4.2	566	15	AA75207	Human zona pelluci
43	86.5	4.2	566	20	AA742479	Human zona pelluci
44	86.5	4.2	566	20	AAW81816	Human ZPB protein.
45	86.5	4.2	566	21	AA782214	Human zona pelluci

ALIGNMENTS

RESULT 1

AA72639

ID AA72639 standard; Protein; 390 AA.

XX AA72639;

AC AA72639;

DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4alpha (GST-4alpha).
DE Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1.

XX Homo sapiens.

OS WO200106015-A1.

XX 25-JAN-2001.

PD 19-JUL-2000; 2000WO-US19741.

XX 20-JUL-1999; 99US-0144694.

XX 13-JUL-2000; 2000US-0593828.

XX (REGC) UNIV CALIFORNIA.

PI Rosen SD, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
DR N-PSDB; AAD02697, AAD02698, AAD02699.
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications
PT Claim 3; Fig 1; 128pp; English.
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
PS alpha). GST-4 gene is found on chromosome 16q23.1.
XX GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
CC Note: The present sequence is also shown in sequence listing (page
CC no: 56) but lacks four nucleotides at its 3' end.
XX
SQ Sequence 390 AA;

Query Match 100.0%; Score 2051; DB 22; Length 390;
Best Local Similarity 100.0%; Pred. No. 2.3e-216;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWLPRESSKTVYVLLLAQTCTLLFIISRPSPAGGEDRVHVLVLSWSGSSFLGQL 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 mwlpfsskvtvlllaqtctllfiisrpgspaggdedrvhvlvlsrsgsflgql 60
QY 61 FSQHPDVFYLMPEAWHVWTTLSQGSAAATLHMAVRLMRSIFLCMDVFDAYMPQSRNLSA 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 fsqhpdvfyilmepawhvwtllsqgsaatlhnavrdlmsiflcmdvfdaympqsrnlsa 120
QY 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTRQPFSLAREACRSYSHVVLKEVRF 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 ffwatralcspacsafrgtiskqdvcktlctrqpfslareacrsyshvvlkevrf 180
QY 181 NLQVLYPLLSDPALNRIHVLRDPRAVLRSRRAAGPILARDNGIVLGTNGKWVEADPHL 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 nlqvlypllsdpalnrihvldrpravlrsraagpillardngivlgtngkwveadphl 240
QY 241 RLIREVCRSHVRIAEAAATLKPPPFLLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQL 300
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 rlirevcrshvriaeatlkpppfllrgyrlvrfedlareplaeiralyaftgltltpql 300
QY 301 EAWIHNIHSGGIGKPIEAFHTSSRNARNVSOAWRHAPFTKILRVQEVCAQALQLGYR 360
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 eawihnihsggigkpieafhtssrnarnvsoawrhalfptkilrvqevcagalgilgyr 360
QY 361 PVYSADQQRDLTLDVLPGRPDHFSWASPD 390
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361 pvysadqqrldtldvlpgrpdhfwaspd 390

RESULT 2
AAB41947
ID AAB41947 standard; Protein; 418 AA.
XX
AC AAB41947;
XX
DT 08-FEB-2001 (first entry)

XX Human OREFX ORF1711 polypeptide sequence SEQ ID NO:3422.
XX
KW Human; open reading frame; OREFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
OS
XX WO200058473-A2.
XX 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US08621.
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
DR N-PSDB; AAC76156.
DR
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
PT
XX Claim 11; Page 2599-2600; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human OREFX open reading frames 1 to 3161. The OREFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an OREFX-associated disorder. The
CC nucleic acids can be used to express OREFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 418 AA;

Query Match 91.0%; Score 1865.5; DB 21; Length 418;
Best Local Similarity 91.0%; Pred. No. 6.3e-196;
Matches 356; Conservative 9; Mismatches 25; Indels 1; Gaps 1;

KW chromosome 8E1.
XX Mus musculus.
OS
XX WO200106015-A1.
PN
XX
PD 25-JAN-2001.
XX 19-JUL-2000; 2000WO-US19741.
PF
XX 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
XX
XX (REGC) UNIV CALIFORNIA.
PA
XX
XX Rosen SD, Lee JK, Hemmerich S;
PI
XX WPI; 2001-138471/14.
DR N-PSDB; AAD02696.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications -
PT
XX
XX Claim 3; Fig 2; 128pp; English.
PS
XX
XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
CC gene is found on chromosome 8E1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
XX
XX Sequence 395 AA;
SQ

Query Match 74.9%; Score 1536.5; DB 22; Length 395;
Best Local Similarity 76.0%; Pred. No. 8.2e-160;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

QY 1 MWLPRFSSKTVTVLLLAQTCLLLFIISRPGSPSSPAGGEDRVHVLVLSWRSGLGQL 60
Db 1 mrlprfstvmlslmvqtg-ilvflvsrqvpspaglgerhvlvlswwrsqsfvgql 59
QY 61 FSQHPDVFYLMPEAWHVTTLTSLQGSAAATLHMAVRDLMRSIFLCMDMDFDAYMPSQSNLSA 120
Db 60 fsqhpdvfylmepawhvwtdtlsqgsapalhmavrdlirsvfclcmdvfdaylpwrrnisd 119
QY 121 FFWNATSRALCSPPPACSAFPRGTISKQDVCKTLCRTQPFSLAREACRSYSHVVLKEVRFF 180
Db 120 lfqwavsralcspvpceafarnisseevckplcatrpfylaqeacssyshvvlkevrf 179
QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAAAGPIIARDNGIVLTNGKWVEADPHL 240
Db 180 nlqvlyplsdpalnlrivilrvdpravlrsregtakalardngivlgtngtwveadprl 239
QY 241 RLIREVCRSHVRIAEAAATLKPFPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPQL 300
Db 240 rvnevcrshvriaeaaalhkpfpfldryrlvryedlarpltvirelyaftgltgtpql 299
QY 301 EAWTNITHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYR 360
Db 300 qtwlnithsgpggarreaafktttsdalsvsqawrhtlpfakirrvqelcggalqllyr 359

QY 361 PVYSADQQORDLTLDLVLPRGPDHFSWAS 388
Db 360 svhselegrdlslldlllprgmdsfkwas 387

RESULT 5
AAM93309
ID AAM93309 standard; Protein; 386 AA.
XX
AC AAM93309;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 2817.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
XX EP1130094-A2.
XX
PD 05-SEP-2001.
XX
XX 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
DR N-PSDB; AAK94229.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PT
XX
PS Claim 8; SEQ ID NO 2817; 1380pp + sequence listing; English.
XX
XX The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequenced of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 386 AA;

Query Match 49.4%; Score 1013; DB 22; Length 386;
Best Local Similarity 52.5%; Pred. No. 2.6e-102;
Matches 207; Conservative 55; Mismatches 104; Indels 28; Gaps 7;

QY 14 LLLAQTTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSWRSGLGQ 59
Db 1 mllpkkmklldflvsqmailalffhmyshnisslsmkaqpermhvvlsswrsqsfvgq 60
QY 60 LFSQHPDVFYLMPEAWHVTTLTSLQGSAAATLHMAVRDLMRSIFLCMDMDFDAYM-PQSRNL 118
Db 61 lfgqhpdvfylmepawhvwmtfkqstawnlhmavrdliravflcdmsvfdamepgprpq 120
QY 119 SAFFNWAATSRALCSPPPACSAFPRGTISKQDVCKTLCRTQPFSLAREACRSYSHVVLKEVR 178

Db	121	sslfqwensralcsapacdiiipqdeiiprahcrlllcsqqpfevvekacrsyshvvlkev	180
QY	179	FFNLQVLYPLSDPALNLRIVHLVDRDPRAVLRSREAAAGPILARDNGIVLGTN-GKWVEAD	237
Db	181	ffnlqslppllkdpnlhlhivlrvdpravrfrsrtkrgdlmidnrivmgqheqklkked	240
QY	238	PHLRLEVCVCRSHVRIAEAAATLKP-PPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTL	296
Db	241	qpvyvmqvicqsqleiyk--tiqslpkalqeryllvryediarapvaqtsrmyefvglef	298
QY	297	TPQLEAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCGALQL	356
Db	299	lphlqtwvhnitrgkmgd--hafhtnardalnvsgawrslpyekvsrlqkacgdamn	356
QY	357	LGYRPVYSADQQRDLTLDLVLPRGPDHFSWASPD	390
Db	357	lgyrhvrsegeqrnlldll-----stwtvpe	383
RESULT	6		
AAAY39918			
ID	AAAY39918	standard; Protein; 386 AA.	
XX			
AC	AAAY39918;		
XX			
DT	08-DEC-1999	(first entry)	
XX			
DE		Human glycosyl sulfotransferase-3 protein sequence.	
XX			
KW		Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;	
KW		selectin binding interaction; inflammation; lymphocyte homing; human;	
KW		secondary lymph organ.	
XX			
OS		Homo sapiens.	
XX			
PN	WO9949018-A1.		
XX			
PD	30-SEP-1999.		
XX			
PF	26-FEB-1999;	99WO-US04316.	
XX			
PR	20-MAR-1998;	98US-0045284.	
PR	12-NOV-1998;	98US-0190911.	
XX			
PA	(REGC) UNIV CALIFORNIA.		
PA	(SYNT) SYNTEx USA INC.		
XX			
PI	Bistrup A, Rosen SD, Tangemann K, Hemmerich S;		
XX			
DR	WPI; 1999-580442/49.		
DR	N-PSDB; AAZ20792.		
XX			
PT	Human and murine glycosyl sulfotransferase 3 and related		
PT	polynucleotides		
XX			
PS	Claim 2; Fig 1; 59pp; English.		
XX			
CC	This sequence is the human glycosyl sulfotransferase-3 (GST-3) of the invention. The nucleic acid sequences, probes and primers derived from these, proteins and antibodies are useful in detecting homologues. The sequences, antibodies and methods are useful in the diagnosis and treatment of diseases associated with selectin binding interactions, including conditions associated with or resulting from the homing of leukocytes to sites of inflammation and the normal homing of lymphocytes to secondary lymph organs.		
XX			
SQ	Sequence	386 AA;	
Query Match			
Best Local Similarity	49.1%;	Score 1008;	DB 20; Length 386;
Matches	206;	Conservative	56; Mismatches 104; Indels 28; Gaps 7;

QY	14	LLLAQTTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSWRSGSSFLGQ	59
Db	1	mlllpkmlkllflvsqmailalfhmyshmisslsmkaqpermhvllsswrsqssfvqg	60
QY	60	LFSQHPDVLYLMEPAWHVWTTLSQSSAATHMAVRDLMRSIFLCMDVFDAYM-PQSRNL	118
Db	61	lfgqhpdvfymepawhvwmtfkqstawmlhmvrdliravflcdmsvdaymegprrq	120
QY	119	SAFFNWAUSTRALCSPPACSAFPRGTISKQDVCKTLCITRQPFSLAREACRSYSHVVLKEVR	178
Db	121	sslfqwensralcsapacdiiipqdeiiprahcrlllcsqqpfevvekacrsyshvvlkev	180
QY	179	FFNLQVLYPLSDPALNLRIVHLVDRDPRAVLRSREAAAGPILARDNGIVLGTN-GKWVEAD	237
Db	181	ffnlqslppllkdpnlhlhivlrvdpravrfrsrtkrgdlmidnrivmgqheqklkked	240
QY	238	PHLRLEVCVCRSHVRIAEAAATLKP-PPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTL	296
Db	241	qpvyvmqvicqsqleiyk--tiqslpkalqeryllvryediarapvaqtsrmyefvglef	298
QY	297	TPQLEAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCGALQL	356
Db	299	lphlqtwvhnitrgkmgd--hafhtnardalnvsgawrslpyekvsrlqkacgdamn	356
QY	357	LGYRPVYSADQQRDLTLDLVLPRGPDHFSWASPD	390
Db	357	lgyrhvrsegeqrnlldll-----stwtvpe	383
RESULT	7		
AAAY79219			
ID	AAAY79219	standard; Protein; 386 AA.	
XX			
AC	AAAY79219;		
XX			
DT	19-JUN-2000	(first entry)	
XX			
DE		Human transferase TRNSFS-11.	
XX			
KW		Transferase; TRNSFS-11; human; antitumour; cell proliferation;	
KW		gastrointestinal disorder; developmental disorder;	
KW		genetic disorder; neurological disorder; reproductive disorder;	
KW		smooth muscle disorder; immunological disorder; inflammation;	
KW		diagnosis; therapy; N-acetylglucosamine 6-O-sulfotransferase.	
XX			
OS		Homo sapiens.	
XX			
FH		Location/Qualifiers	
FT	Key	121	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	107	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	217	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	252	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	364	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	380	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	35	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	50	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	81	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	287	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	243	
FT	Modified-site	/note= "potential O-phosphorylation"	
FT	Modified-site	30	
FT	Modified-site	/note= "potential N-glycosylation"	
FT	Modified-site	308	

```
FT /note= "potential N-glycosylation"
FT 329
FT /note= "potential N-glycosylation"
FT 7..23
FT /note= "transmembrane domain"
XX
PN WO200014251-A2.
XX
PD 16-MAR-2000.
XX
PF 09-SEP-1999; 99WO-US20989.
XX
PR 10-SEP-1998; 98US-0150657.
PR 04-NOV-1998; 98US-0186779.
PR 11-MAY-1999; 99US-0133642.
XX
PA (INCY-) INCYTE PHARM INC.
XX
XX Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
PI Hillman JL, Azimzai Y;
XX
DR WPI; 2000-256996/22.
DR N-PSDB; AAZ94211.
XX
XX Human transferase proteins useful for preventing, diagnosing and
PT treating cancers and developmental, gastrointestinal, genetic,
PT immunological, neurological, reproductive and smooth muscle disorders -
XX
PS Claim 1; Page 90-91; 113pp; English.
XX
XX The present sequence is that of human transferase TRNSFS-11, 1 of
CC 15 claimed novel human transferase proteins of the invention (see
CC AAY79209-23). The sequence was deduced from a cDNA clone (see
CC AAZ94211) isolated from a gallbladder library. It shows homology to
CC mouse N-acetylglucosamine 6-O-sulfotransferase. TRNSFS-11 is
CC expressed in dermatologic and gastrointestinal tissues,
CC especially those associated with inflammation and cell
CC proliferation. The new human transferases and polynucleotides can
CC be used in the diagnosis, prevention and treatment of cancer,
CC developmental disorders, gastrointestinal disorders, genetic
CC disorders, immunological disorders, neurological disorders,
CC reproductive disorders, and smooth muscle disorders. The
CC polypeptides can also be used to raise antibodies, and to screen
CC for agonists and antagonists of transferase activity.
XX
SQ Sequence 386 AA;

Query Match 47.3%; Score 970; DB 21; Length 386;
Best Local Similarity 51.3%; Pred. No. 1.4e-97;
Matches 202; Conservative 52; Mismatches 112; Indels 28; Gaps 7;

QY 14 LLLAQTTCLLFIISRP-----GPSSPAGGEDRVHVLVLSWSRSGSFLGQ 59
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 mllpkkmklillvqsqmailalfhmynhnlsslsmkqaqpermhvvlsswrgssfvqq 60

QY 60 LFSQHPDVLYLMEPAWHVWTTLSQGSAAATLHMAVRDLMRSIFLCMDVFDAYM-PQSRNL 118
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 lfgqhpdvfylmepawhvwmtfkqstawmlhnavrdliravflcdmsvfdaymepgprq 120

QY 119 SAFFNWTATSRALCSPACSAFFRGTTISKQDVCKTLCITRQPFSLAREACRSYSHVVLKEVR 178
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 sslfqwensralcsapacdliipqdesspgltagscavnsplkllcakacrsyshvvlkevr 180

QY 179 FENLQVLYPLSDPALNLRIVHLVRDPRAVLRSRAAGPILARDNGIVLGTN-GKWVEAD 237
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 ffnlqsllypllkdpnlhivhlvrdravfrsrtkrgdmlidsrvmgqheqklkked 240

QY 238 PHRLIREVCRSHVRIAEAAATLKP-PPFLRGRYRLVRFFEDLAREPLAEIRALYAPTCLTL 296
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 qpyvmqvicgsqleiyk--tiqslpkalqeryllvryedlarapvaqtsrmyefvglef 298

QY 297 TPQLEAWIHNITHGSGICKPIEAFHTSSRNARNVSAWRHALPFTKILRVQECAGALQL 356
```

```
Db 299 lphlqtwwhnitrgkmgd--hafhtnardalnvsqarwslpyekvsrlqkacgdamnl 356
| | : : | | | | | | : | | | | | | | | | | : | : | : | | : |
QY 357 LGYRPVYSADQQRDLTLDLVLPRGPDHFSWASPD 390
| | | | | : | : | : | | | : | | : | : |
Db 357 lgyrhvrsegeqrnlldll-----stwtvpe 383
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
AAY39919
ID AAY39919 standard; Protein; 388 AA.
XX
AC AAY39919;
XX
DT 08-DEC-1999 (first entry)
XX
DE Mouse glycosyl sulfotransferase-3 protein sequence.
XX
KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW secondary lymph organ.
XX
OS Mus sp.
XX
PN WO9949018-A1.
XX
PD 30-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04316.
XX
PR 20-MAR-1998; 98US-0045284.
PR 12-NOV-1998; 98US-0190911.
XX
PA (REGC ) UNIV CALIFORNIA.
PA (SYNT ) SYNTEX USA INC.
XX
PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX
DR WPI; 1999-580442/49.
DR N-PSDB; AAZ20793.
XX
XX Human and murine glycosyl sulfotransferase 3 and related
PT polynucleotides
XX
PS Claim 2; Fig 4; 59pp; English.
XX
XX This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of
CC the invention. The nucleic acid sequences, probes and primers derived
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.
XX
SQ Sequence 388 AA;

Query Match 46.7%; Score 958.5; DB 20; Length 388;
Best Local Similarity 55.6%; Pred. No. 2.6e-96;
Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;

QY 42 VHVLVLSWSRSGSFLGOLFQHPDVLYLMEPAWHVWTTLSQGSAAATLHMAVRDLMRSIF 101
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 42 vhlvlswwrgssfvqqlfgqhpdvfylmepawhvwmtftstawkhlhnavrdllrsvf 101

QY 102 LCDMDVFDAYM-PQSRNLSAFFNWTATSRALCSPACSAFFRGTTISKQDVCKTLCITRQPF 160
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 lcdmsvfdaymnpgrkqsslfqweqsralsapvcdfpabeisspkhckllcgqqpfd 161

QY 161 LARACRSYSHVVLKEVRFFENLQVLYPLSDPALNLRIVHLVRDPRAVLRSRAAGPILA 220
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 162 mvekacrsghfvvlkevrflslqalyplltdpsnlhvhvlvrdravfrsrehhttielm 221
```


*	91575	91674:	gap of unknown length
*	91675	93173:	contig of 1499 bp in length
*	93174	93273:	gap of unknown length
*	93274	94320:	contig of 1047 bp in length
*	94321	94420:	gap of unknown length
*	94421	95911:	contig of 1491 bp in length
*	95912	96011:	gap of unknown length
*	96012	97119:	contig of 1108 bp in length
*	97120	97219:	gap of unknown length
*	97220	98221:	contig of 1002 bp in length
*	98222	98321:	gap of unknown length
*	98322	99324:	contig of 1003 bp in length
*	99325	99424:	gap of unknown length
*	99425	100879:	contig of 1455 bp in length
*	100880	100979:	gap of unknown length
*	100980	102554:	contig of 1575 bp in length
*	102555	102654:	gap of unknown length
*	102655	104910:	contig of 2256 bp in length
*	104911	105010:	gap of unknown length
*	105011	107042:	contig of 2032 bp in length
*	107043	107142:	gap of unknown length
*	107143	108339:	contig of 1197 bp in length
*	108340	108439:	gap of unknown length
*	108440	109717:	contig of 1278 bp in length
*	109718	109817:	gap of unknown length
*	109818	111029:	contig of 1212 bp in length
*	111030	111129:	gap of unknown length
*	111130	112643:	contig of 1514 bp in length
*	112644	112743:	gap of unknown length
*	112744	114325:	contig of 1582 bp in length
*	114326	114425:	gap of unknown length
*	114426	115714:	contig of 1289 bp in length
*	115715	115814:	gap of unknown length
*	115815	116991:	contig of 1177 bp in length
*	116992	117091:	gap of unknown length
*	117092	118249:	contig of 1158 bp in length
*	118250	118349:	gap of unknown length

Query Match
Best Local Similarity
Matches 902; Conservative

60.7%;
77.4%;
0;

Score 711.5;
Pred. No. 2.2e-86;
0;

DB 2;
Indels 5;
Gaps 3;

Length 139252;

QY	1	atgtggtgcacggttcttcacgaagacagtgcacagtgcctcctccttgccacagaccacc	60
Db	57972	AUGCGGTACCCCGCTTCTCAGCACACTGTAGTGCTTCGGCTCCTGATAGTGCAG--ACC	58028
QY	61	tgcctcgtctcttcatcatcctcccggccagggccctcatccccagccggcgagggat	120
Db	58029	GGCATCTCTGGTCTTCCTGGNCTCCGGCAAAATGCCATCATCCCCAGACGCCGTGGGGAG	58088
QY	121	cgttgtcacgtgctgtgtctctcgtggtgcgtggcgtgcattcctcttggtggccagctc	180
Db	58089	CATGTGCGATGTGCTGTGTTCTTCCTCCTGGCGCTCGGCTCGTCTTTCGTGGGCCAGCTC	58148
QY	181	ttcagccagcaaccccgacgttcttacctgatggagcccgctggcattgct-ggaccac	239
Db	58149	TTCAGCCAGGCCCGGATGTCTTCTACCTGATGGAGCCAGCTTGGCACCGTGGGGAIGG	58208
QY	240	cctgtcgcagggcagcgcggaacgctgcacatggccgtgcgacactgatcgcttat	299
Db	58209	GTTGTCCGAGGCCAGTAGTGCCTCCGCTCCACATGGCCGTGGCGGACCTGATCCGTTCACT	58268
QY	300	cttttttgcgacatggacgtgttttgatgcctaatgccaacagagccgaacctgtccgc	359
Db	58269	GTTCCTATGCGACATGGACGTTGTTCGATGCCCTACCTACCTGCCCTGGCGCGCAATATCTCGGA	58328
QY	360	ctttttlcaactlggcaacgagcgcgcgtgtgtgtcgcgcgccccctgcagcgcttcc	419
Db	58329	TCTCTTCCAGTGGGCAGGAGCGCGGCGCTGTGTGCTCGCCTCCAGTCTGCGACGCTTCGG	58388
QY	420	ccgaggcaccatcagcaagcaggacgttatgcaagacactgtgcacggcgacgattccag	479
Db	58389	TCGGGGCAACATTAGCAGCGAGGAGTGTGTAAAGCCTCTGTGTGCAACGCGGGCCTTCGG	58448

Fri Sep 27 20:49:20 2002

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaiké,F., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Wleczyk,R., Wooden,S.,
Watlington,S., Williams,G., Williams,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 139252)
Worley,K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15627284.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCIW
Center clone name: CH230-8L17
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 96308 bases at least Q40
Consensus quality: 106296 bases at least Q30
Consensus quality: 113927 bases at least Q20
Estimated insert size: 91054; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 71 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6038: contig of 6038 bp in length
* 6039 6138: gap of unknown length
*
6139
9961
10061
13889
13989
18181
18281
21331
21431
23490
23590
27448
27548
30006
30106
33184
33284
36346
36446
38893
38993
40984
41084
43883
43983
47307
47407
49143
49243
52211
52311
55012
55112
57129
57229
59165
59265
60946
61046
62363
62463
64181
64281
65665
65765
67614
67714
69379
69479
71052
71152
73344
73444
75184
75284
76991
77091
78116
78216
80433
80533
81895
81995
83220
83320
84715
84815
86251
86351
88175
88275
89925
90025
91574: contig of 1550 bp in length
9960: contig of 3822 bp in length
10060: gap of unknown length
13888: contig of 3828 bp in length
13988: gap of unknown length
18180: contig of 4192 bp in length
18280: gap of unknown length
21330: contig of 3050 bp in length
21430: gap of unknown length
23489: contig of 2059 bp in length
23589: gap of unknown length
27447: gap of unknown length
27547: gap of unknown length
30005: contig of 2458 bp in length
30105: gap of unknown length
33183: contig of 3078 bp in length
33283: gap of unknown length
36345: contig of 3062 bp in length
36445: gap of unknown length
38892: contig of 2447 bp in length
38992: gap of unknown length
40983: contig of 1991 bp in length
41083: gap of unknown length
43882: gap of unknown length
47306: contig of 3324 bp in length
47406: gap of unknown length
49142: contig of 1736 bp in length
49242: gap of unknown length
52210: contig of 2968 bp in length
52310: gap of unknown length
55011: contig of 2701 bp in length
55111: gap of unknown length
57128: contig of 2017 bp in length
57228: gap of unknown length
59164: contig of 1936 bp in length
59264: gap of unknown length
60945: contig of 1681 bp in length
61045: gap of unknown length
62362: contig of 1317 bp in length
62462: gap of unknown length
64180: contig of 1718 bp in length
64280: gap of unknown length
65664: contig of 1384 bp in length
65764: gap of unknown length
67613: contig of 1849 bp in length
67713: gap of unknown length
69378: contig of 1665 bp in length
69478: gap of unknown length
71051: contig of 1573 bp in length
71151: gap of unknown length
73343: contig of 2192 bp in length
73443: gap of unknown length
75183: contig of 1740 bp in length
75283: gap of unknown length
76990: contig of 1707 bp in length
77090: gap of unknown length
78115: contig of 1025 bp in length
78215: gap of unknown length
80432: contig of 2217 bp in length
80532: gap of unknown length
81894: contig of 1362 bp in length
81994: gap of unknown length
83219: contig of 1225 bp in length
83319: gap of unknown length
84714: contig of 1395 bp in length
84814: gap of unknown length
86250: contig of 1436 bp in length
86350: gap of unknown length
88174: contig of 1824 bp in length
88274: gap of unknown length
89924: contig of 1650 bp in length
90024: gap of unknown length
91574: contig of 1550 bp in length

/rpt_type=dispersed
complement(462. .771)
/rpt_family="Alu"
/rpt_type=dispersed
772. .945
/rpt_family="L2"
/rpt_type=dispersed
1169. .1338
/rpt_family="MIR"
/rpt_type=dispersed
complement(1553. .1804)
/rpt_family="Alu"
/rpt_type=dispersed
2616. .2916
/rpt_family="Alu"
/rpt_type=dispersed
3309. .3532
/note="similar to EST H12297 (EST cluster Hs.31147)"
3865. .4028
/note="similar to EST H12297 (EST cluster Hs.31147)"
4099. .4406
/rpt_family="Alu"
/rpt_type=dispersed
4793. .5250
/rpt_family="L2"
/rpt_type=dispersed
complement(5251. .5329)
/rpt_family="MIR"
/rpt_type=dispersed
complement(5768. .6080)
/rpt_family="Alu"
/rpt_type=dispersed
complement(6081. .6265)
/rpt_family="Alu"
/rpt_type=dispersed
6111. .6567
/note="similar to EST AA001593 (EST cluster Hs.6853)"
8035. .8350
/rpt_family="Alu"
/rpt_type=dispersed
10545. .11049
/note="similar to EST AI950923 (EST cluster Hs.6853)"
11043. .11195
/rpt_family="Alu"
/rpt_type=dispersed
11200. .11222
/rpt_family="Simple_repeat"
/rpt_type=tandem
11570. .11871
/rpt_family="Alu"
/rpt_type=dispersed
complement(12065. .12198)
/rpt_family="L1"
/rpt_type=dispersed
complement(12205. .12493)
/rpt_family="Alu"
/rpt_type=dispersed
complement(12813. .13110)
/rpt_family="Alu"
/rpt_type=dispersed
join(14185. .14228,16169. .16263,17712. .20821)
/gene="CHST5"
/product="intestinal N-acetylglucosamine
6-O-sulfotransferase"
14185. .20821
/gene="CHST5"
14681. .14713
/rpt_family="Simple_repeat"
/rpt_type=tandem
complement(14715. .14990)
/rpt_family="Alu"
/rpt_type=dispersed
complement(14995. .15288)

repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
complement(15292. .15589)
/rpt_family="Alu"
/rpt_type=dispersed
15668. .15718
/rpt_family="Alu"
/rpt_type=dispersed
15977. .16098
/rpt_family="MIR"
/rpt_type=dispersed
complement(17298. .17433)
/rpt_family="MIR"
/rpt_type=dispersed
complement(18001. .18310)
/rpt_family="Alu"
/rpt_type=dispersed
complement(18420. .18591)
/rpt_family="MIR"
/rpt_type=dispersed
18732. .18767
/rpt_family="Simple_repeat"
/rpt_type=tandem
18967. .20202
/gene="CHST5"
/note="I-GlcNAc6ST; carbohydrate sulfotransferase 5;
expressed in the small intestine and colon"
/codon_start=1
/product="intestinal
N-acetylglucosamine-6-O-sulfotransferase"
/protein_id="AAG26326.1"
/db_xref="GI:11023148"
/translation="MGMRARVPKVAHSTRPPAARMVLPFRSSKTVTVLLAQITCILL
LFIIISRPGPSPPAGGEDRVHVLVLSWSRSGSFLQLESQHPDVFYLMPEPAWHVWITL
SQGSAATLHMAVRDLMSIFLCDMDVFDAYMPQSRNLSAFAFNWATSRALCSPACSAF
PRGTISKQDVCKTLCTRQPTSLAREACRSYSHVVLKVEFFNLQVLYPLLSDPALNLR
IVHLVRDPRAVLRSEAGPILARDNGIVLGTNGKWEADPHLRRLIREVCRSHVRIAE
AATLKPPPPFLRGYRLVRFEDIAREPLAEIRALYAFGLTLTPQLEAWIHNIHSGSGI
GKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCAQALQLLGYRVPVYSADQQQDL
TLDLVLPKRGPDHFSWASPD"
20504. .20806
/rpt_family="Alu"
/rpt_type=dispersed
20869. .20961
/rpt_family="L1"
/rpt_type=dispersed
20962. .21200
/rpt_family="MER4-group"
/rpt_type=dispersed
21325. .21429
/rpt_family="MER4-group"
/rpt_type=dispersed
21430. .21722
/rpt_family="Alu"
/rpt_type=dispersed
21725. .22007
/rpt_family="Alu"
/rpt_type=dispersed
22008. .22040
/rpt_family="Simple_repeat"
/rpt_type=tandem
22041. .22313
/rpt_family="MER4-group"
/rpt_type=dispersed
22325. .22593
/rpt_family="Alu"
/rpt_type=dispersed
22594. .22949
/rpt_family="MER4-group"
/rpt_type=dispersed
23116. .23221
/rpt_family="Simple_repeat"
/rpt_type=tandem

gene /map="16q23.1" /tissue_type="intestine" 1. .2170 /gene="I-GlcNAc-6-ST" 344. .1516 CDS /gene="I-GlcNAc-6-ST" /note="possibly involved in biosynthesis of mucosal ligands for L-selectin" /codon_start=1 /product="N-acetylglucosamine 6-O-sulfotransferase" /protein_id="AAD56000.1" /db_xref="GI:5917706" /translation="MWLPRESKTVTVLLAQTTCLLLFIISRPSPSPAGGEDRVHV LVLSSWRSGSFLGQLFSQHPDVLYLMEPAWHVWTTLSQGSAAATLHMAVRDLMRSIFL CDMDVFDAYMPOSRNLSAFFNWATSRALCSPACSAFPRGTISKQDVCKTLCTRQPF LAREACRSYSHVVLKEVRFNLQVLYPLSDPALNLRIVHLVDPRAVLRSREAAGPI LARDNGIVLGTNGKWEADPHLRLREVCRSHVRIAEEALKPPPPFLRGYRLVRFFD LAREPLAEIRALYAFTGLTLPQLEAWIHNIHSGSIGKPIEAFHTSSRNARNV SQAW RHALPFTKILRVQEVCAQALQLLYGYPVYSADQQORDLTLDLVLPRGPDHFSWASPD" BASE COUNT 449 a 682 c 612 g 427 t ORIGIN

Query Match 100.0%; Score 1173; DB 9; Length 2170; Best Local Similarity 100.0%; Pred. No. 2.7e-147; Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtggctgccacggttctccagcaagacagtgacagtgcctcctcctctggcacagaccacc 60
Db 344 ATGTGGCTGCCACGGTTCTCCAGCAAGACAGTGACAGTGCTCCTCCTGGCACAGACC 403

QY 61 tgcctcctgctcttcaatcatctccggccagggccctcatccccagccggcgaggagat 120
Db 404 TGCCTCCTGCTCTTCATCATCTCCGGCCAGGGCCCTCATCCCCAGCCGGCGGAGGAT 463

QY 121 cgtgtgcacgtgctgtgctgtcctcgtggtgcgtggggtcatccttcttgggcccagctc 180
Db 464 CGTGTGCACGTGCTGCTGTCTCCTCGTGGCGCTCGGGCTCATCCTTCTTGGGCCAGCTC 523

QY 181 ttcagccagcaccgccacgtcttctacctgatggagcccgctggcagtggtggaacacc 240
Db 524 TTCAGCCAGCACCCCGACGCTTCTACCTGATGGAGCCCGCGTGGCATGTGTGGACACC 583

QY 241 ctgtgcagggcagcggcgaacgctgcacatggccgtgcgcacctgatgcgctctatc 300
Db 584 CTGTGCGAGGGCAGCGCGCAACGCTGCACATGSCCGTGGCGACCTGATGCGCTCTATC 643

QY 301 tttttgtgcacatggacgtgtttgatgcctacatgccacagagccgaacacctgtccgcc 360
Db 644 TTTTGTGCGACATGGACGTGTTTGATGCCTACATGCCACAGAGCCGGAACCTGTCCGCC 703

QY 361 tttttcaactgggcaacgagccgcgctgtgctgcgcgcccgccctgcagcccttccc 420
Db 704 TTTTTCACCTGGCAACGAGCCGCGCGCTGTGCTCGCCCGCCGCTGCAGCGCCTTCCC 763

QY 421 cgaggcaccatcagcaagcaggacgtatgcaagacactgtgcacgcggcagccattcagc 480
Db 764 CGAGGCACCATCAGCAAGCAGGACGTATGCAAGACACTGTGCACGCGGCAGCCATTGAGC 823

QY 481 ctggcccgaggacctgcgctcctacagccacgtgtgctcgaaggaggtgcgctcttc 540
Db 824 CTGGCCCGGAGGCGCTGCCGCTCCTACAGCCACGTGGTGTCAAGGAGGTGCGCTTCTTC 883

QY 541 aacctgcaggtgctctaccgctgctcagcgacccccgctcaacctgcgcacgtgcac 600
Db 884 AACTGTCAGGTGCTCTACCCGCTGCTCAGCGACCCCGCGCTCAACCTGCGCATCGTGAC 943

QY 601 ctggtgcgcgaccccgcggtgctgcgtcccggggagggcgggcccgatactggca 660
Db 944 CTGTTGCGCGACCCCGCGGCGGTGCTGCGCTCCCGGGAGGCGGGCCCGGATACTGGCA 1003

QY 661 cgcgacaacggcatcgtgctgggacccaacggcaagtggagggccgacccctcacctg 720

Db 1004 CCGGACAAACGGCATCGTGTGGGCACCAACGGCAAGTGGTGGAGCCGACCCCTCACCTG 1063

QY 721 cgcctgattcgcgaggtgtgcccagccacgtgcatcgccatgcgagggccgcccacactcaag 780
Db 1064 CGCCTGATTGGGAGGTGTGCCGAGCCACCGTGCCTGCGGAGCCGCGCCACACTCAAG 1123

QY 781 ccgccacccttctcgtgcggccgctaccgcctggtgccttcgaggtgacctggcgcggaag 840
Db 1124 CCGCCACCCTTCTCTGCGCGCGCTACCGCCTGCTGCGCTTCGAGGACCTGGCGCGGAG 1183

QY 841 ccgctggcagagatccgcgcaactctacgccttcacccggtccctcaccctcaccgccacgctc 900
Db 1184 CCGCTGGCAGAGATCCGGCACTCTACGCCTTCACCGGCTGACCCCTACGCCACAGCTC 1243

QY 901 gaggcctggatccacaacatcacccacgggtcggggatcgggcaagccaatcgagggcttc 960
Db 1244 GAGCCTGGATCCACAACATCACCCACGGGTGCGGGATCGGCAAGCCAATCGAGGCCTTC 1303

QY 961 catacttcgtctaggaatgcgcgcaacgtctcccagggccttcggccacgcgttgccttc 1020
Db 1304 CATACTTCGTCTAGGAATGCGGCAACGTCTCCAGGCTTCAGGCTGCGGCTGCGCCTTC 1363

QY 1021 actaagatcctgcgcgtgcaggaggtgtgcgcggcgcgctgcagctgctgggtaccgg 1080
Db 1364 ACTAAGATCCTGCGCGTGCAGGAGGTGTGCGCGCGGCTGCAGCTGCTGGGTACCGG 1423

QY 1081 cctgtgtactctgcggaccagcagcagtgacctcacccctggtgctggtgctgccacgaggc 1140
Db 1424 CCTGTGTACTCTGCGGACCAAGCAGCGTGACCTCACCCCTGGATCTGCTGCTGCCACGAGGC 1483

QY 1141 ccagaccacttcagctgggcatcgccctgactga 1173
Db 1484 CCAGACCCTTCAGCTGGGCAATCGCCTGACTGA 1516

RESULT 3
AF246718 3278 bp mRNA linear PRI 31-OCT-2000
LOCUS Homo sapiens intestinal GlcNAc-6-sulfotransferase (CHST5) mRNA,
DEFINITION complete cds, alternatively spliced.
ACCESSION AF246718
VERSION AF246718.1 GI:11055254
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 3278)
AUTHORS Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Ozaki,K., Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., Fujiwara,T., Thonar,E.J., Shimomura,Y., Kinoshita,S., Tanigami,A. and Fukuda,M.N.
TITLE Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene
JOURNAL Nat. Genet. 26 (2), 237-241 (2000)
MEDLINE 20472330
PUBMED 11017086
REFERENCE 2 (bases 1 to 3278)
AUTHORS Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Ozaki,K., Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., Fujiwara,T., Thonar,E.J.-M.A., Shimomura,Y., Kinoshita,S., Tanigami,A. and Fukuda,M.N.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-2000) Glycobiology Program, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
1. .3278
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q22: between D16S512 and D16S515"
1. .3278
/gene="CHST5"
gene

FEATURES	3401 Hillview Avenue, Palo Alto, CA 94304, USA			
	Location/Qualifiers			
source	1. .1462			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/chromosome="16"			
	/map="16q23.1"			
mRNA	/tissue_type="intestine"			
	<229. .>1401			
gene	/gene="I-GlcNAc-6-ST"			
CDS	/product="N-acetylglucosamine 6-O-sulfotransferase"			
	<229. .>1401			
	/gene="I-GlcNAc-6-ST"			
	/gene="I-GlcNAc-6-ST"			
	/note="possibly involved in biosynthesis of mucosal ligands for L-selectin"			
	/codon_start=1			
	/product="N-acetylglucosamine 6-O-sulfotransferase"			
	/protein_id="AAD56001.1"			
	/db_xref="GI:5917708"			
	/translation="MWLPFRSSKTVTVLLLAQTTCLLFIISRPGPSSPAGGEDRVHV LVLSSWRSGSFLQFLFSQHPDVFYLMPEAWHVWTFLLSQGSAATLHMAYRDLMRSLFL CDMDVFDAYMPQSRNLSAFFENWATSRALCSPACSAFPRGTISKQDVCKTLCTRPQFSLAREACRSYSHVVLKEVFNFQLVPLLSDPALNLRIVHLVRDPRAVLRSREAAAGPI LARDNGIVLGTNGKWEADPHLRLIREVCRSHVRIAEAATLKPPFLRGYRLVRFED LAREPLAEIRALYAFTGLTLPQLEAWIHNIHTGSGIGKPIEAFHTSSRNARNVYQAW RHALPFTKILRVQEVACAGALQLLGYRVPVYSADQQRDLTLDLVLPRGPDHFSWASPD"			
BASE COUNT	239 a 510 c 446 g 267 t			
ORIGIN				
	Query Match 100.0%; Score 1173; DB 9; Length 1462;			
	Best Local Similarity 100.0%; Pred. No. 3e-147;			
	Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 atgtgtgctgccacggttctccagcaagacagtgacagtgctcctcctcctggcacagaccacc	60		
Db				
QY	229 ATGTGGCTGCCACGGTTCTCCAGCAAGACAGTGACAGTGTCTCTCTCTGGCACAGACCACC	288		
Db				
QY	61 tgcctcctgtcttctcatcatctcccgccaggccctcatcccgccggcgaggat	120		
Db				
QY	289 TGCCTCTGCTCTTCATCATCTCCGGGCCAGGGCCCTCATCCCGACCGCGCGGAGGAT	348		
Db				
QY	121 cgtgtgcacgtgctggtgctgtcctgtggcgctcggtgcctccttcttgggcccagctc	180		
Db				
QY	349 CGTGTGCACGTGCTGGTGTCTCTCTGCTGGCGCTCGGGCTCATCTCTTGGGCCAGCTC	408		
Db				
QY	181 ttcagcagcagccccgacgtctcttacctgatggagccccgctgggcatgtgtggaccacc	240		
Db				
QY	409 TTCAGCCAGCACCCCGACGCTCTTACCTGATGGAGCCCCGGTGGCATGTGTGGACCACC	468		
Db				
QY	241 ctgtgcagggcagcgcggcaacgctgcacatggcogtgcggacctgatgcgtctatc	300		
Db				
QY	469 CTGTGCAGGGCAGCGCGGCACCGCTGCACATGGCCGTGCGCGACCTGAIGCGCTCTATC	528		
Db				
QY	301 tttttgtgcacatggacgtgtttgatgcctacatgccacagagccgaaacctgtccgcc	360		
Db				
QY	529 TTTTGTGCGACATGGACGTGTTGATGCCCTACATGCCACAGAGCCGAAACCTGTCCGCC	588		
Db				
QY	361 tttttcaactgggcaacgagccgcgtgtgtcctgcgcgcgcctgcagcgcccttccc	420		
Db				
QY	589 TTTTCAACTGGGCAACGAGCGCGCTGTGCTCGCGCCCGCCTGCAGCGCCTTTCCC	648		
Db				
QY	421 cgaggcaccatcagcaagcaggacgtatgcaagacactgtgcacgcggcagccattcagc	480		
Db				
QY	649 CGAGCACCATCAGCAAGCAGGACGTAIGCAAGACACTGTGTCACGCGGCAGCCATTTCAGC	708		
Db				
QY	481 ctggcccgaggagccctgcgcctcctacagccacgtgtgtcctcaaggaggtgcgtcttc	540		
Db				
QY	709 CTGGCCCCGGAGGCCTGCCGCTCCTACAGCCACGTGGTGTCTCAAGGAGGTGCGCTTCTTC	768		
Db				
QY	541 aacctgcaggtgctctatacccgctgctcagcgacccccgcgtcaacctgcgcac	600		

Db	769	AACCTGCAGGIGCTCTACCCGCTGCTCAGCGACCCCGCGCTCAACCTGGCATCGTGAC	828
Qy	601	ctggtgcgcgacccggggccgtgctgcgtcccgggaggcgggcccgcatactggca	660
Db	829	CTGGTGGCGACCCGGGGCCGTGCTGCGCTCCCGGGAGGCGGGCCCGATACTGGCA	888
Qy	661	cgcgacaacggcatcgtgctggggcaccacaacggcaagtgggtggaggccgacctcacctg	720
Db	889	CGCGACAACGGCATCGTGTGGGCACCAACGGCAAGTGGTGGAGGCCGACCCCTCACCTG	948
Qy	721	cgcctgattcgcgaggtgtgccgcagccacgtgcgcacatcgccgagggccgccaactcaag	780
Db	949	CGCCTGATTCGCGAGGTGTGCCGACGCCACGTGCGCATGCGCGAGGCCGCCACTCAAG	1008
Qy	781	ccgcccaccttctcgcggccgctaccgcctgctgcgttcgaggaacctgcccgggag	840
Db	1009	CCGCCACCCCTCTCTGCGGGCCGCTACCGCCTGGTGGCTTCGAGGACCTGGCGGGGAG	1068
Qy	841	ccgctggcagagatccgcgcactctacgccttcacggcctgacctcacccacacagctc	900
Db	1069	CCGTGGCAGAGATCCGGCACTCTACGCCCTTCAACGGCCTGACCCCTCACGCCACAGCTC	1128
Qy	901	gaggcctgcatccacaacatcacccacgggtcggggatcggaagccaatcgaggccttc	960
Db	1129	GAGGCTGGATCCACAACATCACCCACGGTCCGGGATCGGCAAGCCAATCGAGGCCCTTC	1188
Qy	961	catacttcgttaggaatcgcgcgaacgtctcccaggcctggcgccacgcgttgcccttc	1020
Db	1189	CATACTTCGTCTAGGAATCGCGCAACGTCTCCAGGCCCTGGCGCCACGCGTTGCCCTTC	1248
Qy	1021	actaagatccctgcgctgcaggaggtgtgcgcggcgctgcagctgctgggctaccgg	1080
Db	1249	ACTAAGATCCTGCGCGTGCAGGAGGTGTGCGCGCGCGCTGCAGCTGCTGGGCTACCGG	1308
Qy	1081	cctgttactctcggaccagcagcgtgacctcacctgacctggtgctgctgccacgagc	1140
Db	1309	CCTGTGTACTCTGCGGACCAGCAGCGTGACCTACCCCTGGATCTGGTGTGCCACGAGGC	1368
Qy	1141	ccagaccacttcagctgggcatcgccctgactga	1173
Db	1369	CCAGACCACCTCAGCTGGGCATCGCCTGACTGA	1401
RESULT 2			
AF176838			
LOCUS			
DEFINITION Homo sapiens intestine N'-acetylglucosamine 6-O-sulfotransferase (I-GlcNAc-6-ST) mRNA, complete cds.			
ACCESSION AF176838			
VERSION AF176838.1 GI:5917705			
KEYWORDS human.			
SOURCE Homo sapiens			
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE 1 (bases 1 to 2170)			
AUTHORS Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.			
TITLE Cloning and characterization of a mammalian N-acetylglucosamine-6-sulfotransferase that is highly restricted to intestinal tissue			
JOURNAL Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)			
MEDLINE 99423499			
REFERENCE 2 (bases 1 to 2170)			
AUTHORS Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.			
TITLE Direct Submission			
JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience, 3401 Hillview Avenue, Palo Alto, CA 94304, USA			
FEATURES			
source			
1. .2170			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/chromosome="16"			

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 14:16:59 ; Search time 15810.8 Seconds
(without alignments)
1552.539 Million cell updates/sec

Title: US-09-593-828-4
Perfect score: 1173
Sequence: 1 atgtgggtgccacggttctc.....gctgggcacgcctgactga 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:**

- 1: gb_ba:**
- 2: gb_htg:**
- 3: gb_in:**
- 4: gb_om:**
- 5: gb_ov:**
- 6: gb_pat:**
- 7: gb_ph:**
- 8: gb_pl:**
- 9: gb_pr:**
- 10: gb_ro:**
- 11: gb_sts:**
- 12: gb_sy:**
- 13: gb_un:**
- 14: gb_vi:**
- 15: em_ba:**
- 16: em_fun:**
- 17: em_hum:**
- 18: em_in:**
- 19: em_mu:**
- 20: em_om:**
- 21: em_or:**
- 22: em_ov:**
- 23: em_pat:**
- 24: em_ph:**
- 25: em_pl:**
- 26: em_ro:**
- 27: em_sts:**
- 28: em_un:**
- 29: em_vi:**
- 30: em_htg_hum:**
- 31: em_htg_inv:**
- 32: em_htg_other:**
- 33: em_htgo_inv:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%			
Result No.	Score	Query Match Length	Description

1	1173	100.0	1462	9	AF176839	Homo sapi
2	1173	100.0	2170	9	AF176838	Homo sapi
3	1173	100.0	3278	9	AF246718	Homo sapi
4	1173	100.0	71503	9	AF219991	Homo sapi
5	1173	100.0	157337	9	AC009163	Homo sapi
6	1173	100.0	159072	2	AC026419	Homo sapi
7	1173	100.0	194832	9	AC025287	Homo sapi
8	924.6	78.8	1647	6	AX327330	Sequence
9	924.6	78.8	2544	9	AF219990	Homo sapi
10	924.6	78.8	3786	9	AF280086	Homo sapi
11	924.6	78.8	159072	2	AC026419	Homo sapi
12	924.6	78.8	208185	2	AC009105	Homo sapi
13	745.2	63.5	1740	10	AF176841	Mus muscu
14	745.2	63.5	1989	10	AF176840	Mus muscu
15	711.6	60.7	139252	2	AC095664	Rattus no
16	400	34.1	2032	9	AF131235	Homo sapi
17	398.4	34.0	1333	9	AF149783	Homo sapi
18	398.4	34.0	1992	9	AF280088	Homo sapi
19	398.4	34.0	2011	9	AK026635	Homo sapi
20	398.4	34.0	183228	9	AC010547	Homo sapi
21	358.8	30.6	2201	10	AF109155	Mus muscu
22	354	30.2	1926	10	AF131236	Mus muscu
23	235.2	20.1	2156	6	AR071396	Sequence
24	235.2	20.1	2156	6	E14937	Human mRNA
25	235.2	20.1	2731	9	AB012192	Homo sapi
26	235.2	20.1	6961	9	AB017915	Homo sapi
27	235.2	20.1	178251	9	AC073370	Homo sapi
28	235.2	20.1	196465	9	AC022392	Homo sapi
29	196.8	16.8	2354	5	CHKC6ST	Chicken mRN
30	196.8	16.8	2354	6	AR050577	Sequence
31	196.8	16.8	2354	6	E12369	DNA encodin
32	196.2	16.7	90256	9	AP001582	Homo sapi
33	196.2	16.7	137499	2	AC015931	Homo sapi
34	196.2	16.7	174187	2	AP002792	Homo sapi
35	196.2	16.7	215647	2	AC068591	Homo sapi
36	190.6	16.2	1425	10	AF178689	Rattus no
37	181	15.4	1654	10	AB008937	Mus muscu
38	181	15.4	7515	10	AB062107S3	Mus muscu
39	153.8	13.1	2227	9	AF083066	Homo sapi
40	153.8	13.1	3029	9	AK023268	Homo sapi
41	153.8	13.1	3535	9	AB021124	Homo sapi
42	153.8	13.1	4862	9	AB021125	Homo sapi
43	153.8	13.1	168056	2	AC055737	Homo sapi
44	153.8	13.1	191644	2	AC018450	Homo sapi
45	148.8	12.7	2393	6	AX321227	Sequence

ALIGNMENTS

RESULT	1
AF176839	AF176839
LOCUS	1462 bp DNA linear PRI 22-SEP-1999
DEFINITION	Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase (I-GlcNAc-6-ST) gene, complete cds.
ACCESSION	AF176839
VERSION	AF176839.1 GI:5917707
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 1462)
TITLE	Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S. Cloning and characterization of a mammalian N-acetylglucosamine-6-sulfotransferase that is highly restricted to intestinal tissue
JOURNAL	Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)
MEDLINE	99423499
REFERENCE	2 (bases 1 to 1462)
AUTHORS	Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
TITLE	Direct Submission
JOURNAL	Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,

```
FEATURES                               Location/Qualifiers
source                                1..429
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F930006H10"
/clone_lib="RIKEN full-length enriched, adult inner ear"
/tissue_type="inner ear"
/dev_stage="adult"

BASE COUNT      63 a   138 c   128 g   100 t
ORIGIN

Query Match      20.4%; Score 239.8; DB 9; Length 429;
Best Local Similarity 77.2%; Pred. NO. 4e-32;
Matches 305; Conservative 0; Mismatches 87; Indels 3; Gaps 1;

QY 1 atgtggctgccacgggtctccagcaagacagtgcacagtctcctcctggcacagaccacc 60
   ||| ||| || || || || || || || || || || || || || || || || || ||
Db 38 ATGCGGCTACCCCGTTCTCCAGCACTGTCTCGCTTCCTGCTCGCTGCTGCTGCTGCTGCT 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 tgcctcctgctcttcctcatctctcccgccagggccctcatctcccccagccggcgaggat 120
   || |||| |||| || |||| || || || || || || || || || || || || || ||
Db 95 GGCAFCCTGGTCTCCTGGTCTCCCGGGAAGTGCCATCGTCCCCACGAGCCCTTGGGGAG 154
   || |||| |||| || |||| || || || || || || || || || || || || || ||

QY 121 cgtgtgcacgtgctgggtgctgtctcctcgtggcgctcgtggcgtcctccttcttgggccagctc 180
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 155 CGTGTGCACGTGTGTGTACTGTCTCCTCGTGGCGCTCGGGCTCGTCTTCGTGGGCCAGCTC 214
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 181 ttcagccagcaccccgacgtcttctacctgatggagcccgctggcagtgtgtggaccacc 240
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 215 TTCAGCCACACCCCGATGTCTTCTACCTGATGGAGCCGGCTTGGCACGTCCTGGGATACG 274
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 241 ctgtgcaggcgagcgcggaacgctgcacatggccgtgcgacacctgatgcgctctatc 300
   |||| |||| |||| || || || || || || || || || || || || || || || || ||
Db 275 TTGTGCGCAGGGCAGTGCCCCCGCACTCAACATGGCCGTGGGTGACCTGATCAGCTCAGTG 334
   || || || || || || || || || || || || || || || || || || || || || ||

QY 301 tttttgtcgacatggacgtgtttgatgcctacatgccacagagccgaaacctgtccgcc 360
   || || || || || || || || || || || || || || || || || || || || || ||
Db 335 TTCCTATGCGACATGGACGATATTGATGCCTACCTGCCCTGGCGCCCGAACAACATCTCGGAT 394
   || || || || || || || || || || || || || || || || || || || || || ||

QY 361 tttttcaactgggcaacgagccgcgcgctgtgtctc 395
   || || || || || || || || || || || || || || || || || || || || || ||
Db 395 CTCTTCCAGTGGCGGTGAGCCGCGCATTTGTGCTC 429
```


340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National Research Initiative, Animal Genome Resource Grant AG 99-3205-8534 to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED from Washington University Genome Center. Vector Trimmi g: Cross-match from Washington University Genome Center PHRAP suite. This sequence is vector free and at least 200 bp in length. PCR primers

FORWARD: TAAACGACTCACTATAGGG

BACKWARD: ATTAACCTCACTAAAG

Insert Length: 517 Std Error: 0.00

Plate: BP250022A10 row: E column: 10

Seq primer: AGCGGATAACAATTTCACACAGGA

High quality sequence stop: 517.

FEATURES
source

```
Location/Qualifiers
1. .517
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="BP250022A10E10"
/clone_lib="Soares normalized bovine placenta"
/ssx="female"
```

BASE COUNT	70 a	177 c	170 g	100 t
ORIGIN				

Query Match	23.1%	Score 271.2;	DB 10;	Length 517;
Best Local Similarity	82.1%	Pred. No. 1.5e-37;		
Matches 325;	Conservative	0;	Mismatches 68;	Indels 3;
				Gaps 1;

Qy 1 atgtggtgccacggttctccagcaagacagtacagtgctcctcctggtcacagaccacc 60
||||| ||| ||||||| + ||||| ||||| ||||| |||||
Db 125 ATGTGGTGGGGCGGTCTCCAGCACGGGGGTGACTGGCTCCTGCTGGCGCAGACGC -- 182

Qy 61 tgctctgctcttcatcatctcccgggccaggggccctcatccccagccggcgggagagat 120
 |||||
 Db 183 -GCCCTGCTCTTCTCTGATCTCCCGGGCCACAGACTGCCGCTTGGGGCCAGCAGCAGGAG 241
 |||||

QY 121 cgtgtgcaagtctgggtgtgtctcctcgtagcgctgggcgcacatccttcttgaggccagctc 180
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 CGGGTCATGTGCTAGTGTCTCCTCGTGGCGCTCGGCCTCGTCTCGTGGGCACGCTC 301

[illegible]

QY 241 ctgtcgagggcagcgcgcaacgctgcacatggccgtgcgcgacctgatgcgctctatc 300
 |||||
Db 362 CTGTGCGAGGGCAGCGCCCTTGGCGCTGCACATGGCAGTCGCCGACCTGGTGCGCTCAGTC 421

[illegible]

QY	361	tttttcaactgggcaacgagcggcgctgtgctcg	396
Db	482	CTCTTCCAATGGGCGGAGAGCGGGCTCTGTGCTCG	517

```

RESULT 13
BI824736
LOCUS      BI824736
DEFINITION 603033615F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5174955 5',
            linear mRNA 260 bp
            EST 04-OCT-2001
            mrna clone IMAGE:5174955 5',
            mRNA sequence.

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 260)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg. ph.D.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Plate: LLAML1435 row: m column: 04
High quality sequence stop: 260.

FEATURES
SOURCE

Location/Qualifiers
1. ,260

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="IMAGE:5174955"  
/clone_lib="NIH_MGC_115"  
/lab_host="DH10B"
```

/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH MGC Library."

BASE COUNT	38 a	103 c	82 g	37 t
ORIGIN				

Query Match	22.2%;	Score 260;	DB 10;	Length 260;
Best Local Similarity	100.0%;	Pred. No. 1.2e-35;		
Matches 260;	Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;

537 cttaaactgcaggtgctctacccgctgctcagcgaccccgcgctcaacctgcgcatcgt 596
 Qy |||||
 Db 1 ctttaacctgcagggtgctctacccgctgctcagcgaccccgcgctcaacctgcgcatcgt 60

QY 597 gcaactggtgcgcgaacccggcgccctgctgcgtcccggaggcggcgccgatact 656
|||||
Ddb 61 GCACCTGGTGC GGACCCGCGGCCGTGCTGGCTCCCGGGAGCGCGGCCGATACT 120
|||||

QY 657 ggaacgagacaacggcatcgtgctggggcaccaacggaagtgggtggaggcgacccctca 716
 |||||
 Db 121 GGCACGGGACAAACGGCATCGTGTCTGGGCACCAACGGCAAGTGGTGGAGCGCACCCCTCA 180
 |||||

Qy	717	cctgcgcctgattcgcgaggtgtgccgcagccacgtggcatcgccgaggcgccacct	776
D _b	181	cctgcgcctgatTCGCGAGGTGTGCCGAGCCACGTGGCATCGCCGAGGCGCCCACT	240

Qy 777 caagcgcgcaccccttctgc 796
|||||
Db 241 CAAGCGGCACCCCTTCCTGC 260

RESULT	14
AZ405100/c	
LOCUS	AZ405100
DEFINITION	1M0173I08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0173I08 R, DNA sequence.
	433 bp
	DNA linear
	GSS 03-OCT-2000

```
Db 498 ATCCACAAAGATGAAATCATCCCGGGGCTCACTGCAGGCTCCTGTGTCAGTCAACAGCCC 557
QY 475 ttacgcctggccggaggcctgcccgtcctacagccacgtggtgctcaaggaggtgcgc 534
Db 558 TTTCAGGTGGTGAGAGGGCTGCCGCTCCTACAGCCACGTGGTGTCTAAGGAGGTGCGC 617
QY 535 ttcttaacctgcaggtgctctaccgcgtgctcagcagcaccgcgcctcaacctgcgcac 594
Db 618 TTCTTAACCTGCAGTCCCTTACCCGCTGCTGAAAGAGACCCCTCCCTCAACCTGCATATC 677
QY 595 gtgcacc-tggtgcgcgacccggggccgtgctgcgtcccgga 638
Db 678 GTGCACCTTGGTCCGGGACCCCGGCGGTGTTCGGTTCCCGAGA 722

RESULT 11
CNS04QFN/c
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
129006 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL302540
VERSION
AL302540.1 GI:8181872
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE
1 (bases 1 to 849)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 849)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
JOURNAL
REFERENCE
3 (bases 1 to 849)
Genoscope.
Direct Submission
JOURNAL
COMMENT
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
Location/Qualifiers
1..849
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="129006"
/clone_lib="G"
/note="Genoscope sequence ID : COBG129BH03SP1-end :
PUC-Ori"
BASE COUNT 137 a 248 c 273 g 184 t 7 others
ORIGIN
Query Match 23.7%; Score 278; DB 12; Length 849;
Best Local Similarity 61.0%; Pred. No. 1.1e-38;
Matches 478; Conservative 3; Mismatches 299; Indels 3; Gaps 2;

QY 385 gcgctgtctgcgcgcgcgcgcgcgccttccccaggccaccatcagcaagcaggac 444
Db 824 GGGCTGTGCTCGCCCGCCGCTGTCTCTCACCCCGCGGAGGATGAGCGATCAGCCC 765
QY 445 gtatgcaagacactgtgcacgcgcgcgcgcgccttcagcctggcccgaggcctgccgctcc 504
```

```
Db 764 GAGTGCITGAAAAAATGCGTGCCTGCAAGGGTCTGCACGGGCTAAGAACGCATCGACACC 705
QY 505 tacagccacgtggtgctcaaggaggtgcgtcttcttaacctgcaggtgctctacccgctg 564
Db 704 TACAGCCACGTGGTGTGAAATCCCGTGGCTCTTCGAGCTGGAGTCCCTGTATCCGCTC 645
QY 565 ctcagcgacccgcgcgtcaacctgcgcacgtgcacacctggtgcgcga-ccgcgggcccgt 623
Db 644 CTGCAGGACCCCAACCTGAACCTCGGCATCGTCCACCTGGTCCGGACCCCTCGGGCCGT 585
QY 624 gctgcgtcccgaggcgcgccggcccgatactggcacgcgacaacggcatcgtgctggg 683
Db 584 RGMAAAAWCMAGAGAACAGTCGGCCNAAACGCTTCGTGCG--CGACGACGCCGTCACTT 527
QY 584 caccacgcgcaagtgggtggaggccgacccctcacctgcgcctgattcgcgaggtgtgccg 743
Db 526 GGAGAACAGGCCCATGCCGGCCCGAGGTTGCTGTATCAGGTCAATGCAGAGWTCTGCCG 467
QY 744 cagccacgtgcgcacatgcgcgagggcgccacactcaagcgcacaccttctgcgcggccg 803
Db 466 CAGCCACGTGCCCATCAGCGAGAGGGCCACGCTGAAGGCCCGCCTTTTCTCAAAGGCCG 407
QY 804 ctaccgcctggtgcgttcgaggaacctggcggggagcggtggcagagatccgcgcact 863
Db 406 CTACAAAATGTGCTCGCTTCGAGGACCTGGTGGGGACCCCGCTCGGGAGATCGGCGCAT 347
QY 864 ctacgccttcacgcgcctgacctcacgcacagctcgaggcctggatccacaacatcac 923
Db 346 CTACAAGTTCGTGCGTCTGGAGATGACCAAGCAGCAGTGGGCCAGTGGATTACCAAGGTAAC 287
QY 924 ccacgggtcggggatcggaagccaatcgaggccttccatacttgcgtctaggaatgcgcg 983
Db 286 CCACGGCAAAGGCCCGCCACCAAGAAGGACGCTTCGCCATCACCTCCAGAGACGCCGC 227
QY 984 caacgtctcccaggcctggcgccacgcgttgcccttcaactaagatcctgcgcgtgcagga 1043
Db 226 CGACGCTCTCCAGGCTTGGCGTTCCGGCTTCCGCCACCAAGGTCAACCGCATCCAGGA 167
QY 1044 ggtgtgcgcgcgcgcgcgtgcagctgctgggtacccgacctgtgtactctgcggaccagca 1103
Db 166 AGTGTGCCGGGGGCCCATGACGCTGTTGGGATACAGGACCCGTGGCAGTGAAGAACA 107
QY 1104 gcgtgacctcacctggatctggtgctgcacgagggccagaccacttcagctgggcatc 1163
Db 106 GAAGAAGCTAAGCGTCGACCTGCTGGTGCCCGCCAGGAACCATACAGGTTCTCTGTTACC 47
QY 1164 gcc 1166
Db 46 GTC 44

RESULT 12
BF042384
LOCUS
DEFINITION
BP250022A10E10 Soares normalized bovine placenta Bos taurus cDNA
clone BP250022A10E10 5', mRNA sequence.
ACCESSION
BF042384
VERSION
BF042384.1 GI:10759439
KEYWORDS
EST.
SOURCE
Bos taurus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 517)
Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
,J.H.
TITLE
Bovine ESTs
JOURNAL
COMMENT
Unpublished (2000)
Contact: Lewin, H. A.
W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
```


QY 443 acgtatgcaagacactgtgcacgcgggcagccattcagcctgccccgggaggccctgcgcgt 502
|||||
Db 376 ACGTATGCAAGACACTGTGCACGGCGGCACTTCAGCCTGGCCCGGAGGCGCTGCCGCT 317

QY 503 cctacagccacgtggtctcaaggagggtgcgcttttcaacctgcaggtgctctaccgc 562
|||||
Db 316 CCTACAGCCACGTTGGTCTCAAGGAGGTGGCGTTCTTCAACCTGCAGGTGCTCTACCGGC 257

QY 563 tgctcagcgaaccccgctcaacctgcgcctcgtgcacactggtgcgcgaccccgcgggccg 622
|||||
Db 256 TGCTCAGCGACCCCGCGCTCAACCTGCGCATCGTGCACTGGTGCGCGACCCCGCGGCGCG 197

QY 623 tgctgcgtcccgaggcgcgccgatactggaacgcgcgacaaacggcctcgtgtgg 682
|||||
Db 196 TGCTGCGCTCCCGGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 137

QY 683 gcaccaacggcaagtgggtggaggccgacccctcactgcgcctgattcgcgaggtgtgcc 742
|||||
Db 136 GCACCAACGGCAAGTGGGTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 77

QY 743 gcagccacgtgcgcctgcgcgagcgccgacactcaagcgccaccttccctgcgcgcc 802
|||||
Db 76 GCAACCACGTGCGGATTGCCGAAGCCCAACACACTCAAAACCGCCACCCCTTTTGGGGGGGC 17

QY 803 gcta 806
| |
Db 16 NGGA 13

RESULT 3
BF197521/c
LOCUS
DEFINITION
7084a08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3642903 3'
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
6-O-SULFOTRANSFERASE. ;, mRNA sequence.

ACCESSION
BF197521
VERSION
BF197521.1 GI:11086670
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 525)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
High quality sequence stop: 451.
Location/Qualifiers
1..525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3642903"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library

(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo. "
BASE COUNT 90 a 163 c 187 g 85 t
ORIGIN

Query Match 44.8%; Score 525; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 2.4e-81;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 catgtgtggaccaccctgtgcagggcagcgcggaacgctgcacatggccgtgcgcgac 285
|||||
Db 525 CATGTGTGGACCAACCCCTGTGCGAGGGCAGCGCGGAACGGTGCACATGGCCGTGCGCGAC 466

QY 286 ctgatgcgtctatatctttttgtgcacatggacgtgtttgatgcctacatgccacagagc 345
|||||
Db 465 CTGATGCGCTCTATCTTTTGTGCGACATGGACGTGTTTGATGCCTACATGCCACAGAGC 406

QY 346 cgaacctgtccgcctttttcaactgggcaacgagccgcgcgtgtgctgcgcgcgcc 405
|||||
Db 405 CGAAACCTGTCCGCCCTTTTCAACTGGGCAACGAGCCGCCGTGTGCTCGCCGCCGCC 346

QY 406 tcgagcgctttccccgaggcaccatcagcaacgaggacgtatgcaagacacactgtgcacg 465
|||||
Db 345 TGCAGCGCCTTTCCCGGAGGCACCATCAGCAAGCAGGACGTATGCAAGACACTGTGCACG 286

QY 466 cggcagccattcagcctggccccggaggcctgcgcctcctacagccacgtggtgctcaag 525
|||||
Db 285 CGGCAGCCATTTCAGCCTGGCCCGGAGGCGCTGCCGCTCCTACAGCCACGTGGTGTCAAG 226

QY 526 gagtgcgcttcttcaacctgcaggtgctctacccgcgtgcacgacccccgcgtcaac 585
|||||
Db 225 GAGGTGCGCTCTTCAACCTGCAGGTGCTCTACCCGCTGCTCAGCGACCCCGCGCTCAAC 166

QY 586 ctgcgcctcgtgcacactggtgcgcgacccgcgcgcgtgctgcgtcccgaggcgcg 645
|||||
Db 165 CTGCGCATCGTGCACCTGTTGGCGGACCCCGGGCGGTGCTGGCGTCCCGGAGGCGGCG 106

QY 646 ggcccgatactggcagcgcaacggcctcgtgcaggtgtgcccagccac 705
|||||
Db 105 GGCCCGATACTGGCACGCGACAAACGGCATCGTGTGGCCACCAACGGCAAGTGGTGGAG 46

QY 706 gccgacctcacctgcgcctgattcgcaggtgtgcccagccac 750
|||||
Db 45 GCCGACCCCTCACCTGGCCTGATTTCGCGAGGTGTGCCGAGCCAC 1

RESULT 4
AI824100/c
LOCUS
DEFINITION
AI824100 620 bp mRNA linear EST 21-DEC-1999
WJ46C01.X1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405856 3'
similar to TR:O75667 O75667 DJ71L16.4 ;contains PTR5.b2 PTR5
repetitive element ;, mRNA sequence.

ACCESSION
AI824100
VERSION
AI824100.1 GI:5444771
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 620)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
High quality sequence stop: 451.
Location/Qualifiers
1..525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3642903"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 09:23:53 ; Search time 10470.7 Seconds
(without alignments)
1512.025 Million cell updates/sec

Title: US-09-593-828-4
Perfect score: 1173
Sequence: 1 atgtggctgccacggttctc.....gctgggcacgcctgactga 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	657.4	56.0	695	10 BE857538	BE857538 7g01a08.x
C 2	631	53.8	735	10 BE858652	BE858652 7g01a09.x
C 3	525	44.8	525	10 BF197521	BF197521 7o84a08.x
C 4	487.6	41.6	620	9 AI824100	AI824100 wj46c01.x
C 5	420.8	35.9	436	9 AW081348	AW081348 xc41b06.x
C 6	352.4	30.0	1923	11 AK009113	AK009113 Mus muscu
C 7	346.4	29.5	1067	12 CNS03KN7	AL248380 Tetraodon
C 8	310.6	26.5	657	12 AG035205	AG035205 Pan trogl
C 9	290.6	24.8	568	9 AI115260	AI115260 ui43c07.y
C 10	284.2	24.2	954	10 BI823850	BI823850 603039012
C 11	278	23.7	849	12 CNS04QFN	AL302540 Tetraodon
C 12	271.2	23.1	517	10 BF042384	BF042384 BP250022A
C 13	260	22.2	260	10 BI824736	BI824736 603033615
C 14	240.2	20.5	433	12 AZ405100	AZ405100 1M0173I08
C 15	239.8	20.4	429	9 BB849113	BB849113 BB849113
C 16	224.4	19.1	695	10 BG964671	BG964671 602831875
C 17	224.4	19.1	783	10 BG963298	BG963298 602827716

18	218.8	18.7	852	10 BG966340	BG966340 602832826
19	174.6	14.9	571	10 BM129080	BM129080 if17c04.y
20	159.4	13.6	965	10 BF579746	BF579746 602095056
21	157.6	13.4	599	10 BJ031352	BJ031352 BJ031352
22	148.6	12.7	322	9 AII56825	AII56825 ui44c08.y
23	146.6	12.5	536	10 BM245312	BM245312 K0722H07-
24	134.6	11.5	2070	11 AK011202	AK011202 Mus muscu
25	134.4	11.5	793	10 BI102274	BI102274 602885587
26	131.8	11.2	497	10 BM246681	BM246681 K0741E04-
27	130.6	11.1	955	10 BG107354	BG107354 602290740
28	129.2	11.0	634	9 AU180328	AU180328 AU180328
29	116.2	9.9	662	10 BF344303	BF344303 602017370
30	111.2	9.5	1070	10 BF163765	BF163765 601769868
31	109.2	9.3	540	9 AW412223	AW412223 uq47g02.y
32	103	8.8	103	9 AI282873	AI282873 qt87e06.x
33	103	8.8	297	9 AA261202	AA261202 va49a06.r
34	103	8.8	500	9 AI529474	AI529474 va49a06.y
C 35	99.8	8.5	135	9 AI824198	AI824198 wj36a01.x
36	97.4	8.3	494	9 AI088880	AI088880 qa17a07.x
37	97.4	8.3	539	10 BF593996	BF593996 nac20a08.
38	97.4	8.3	604	10 BE857485	BE857485 7f98g02.x
39	97.4	8.3	722	9 AW027448	AW027448 wv73h09.x
40	96.6	8.2	714	10 BI115837	BI115837 602866294
41	96.4	8.2	675	9 AI939595	AI939595 tf88e11.x
42	96.2	8.2	564	10 BM488009	BM488009 pgm2n.pk0
43	94.8	8.1	828	10 BI665990	BI665990 603287043
44	94.2	8.0	468	12 AQ939450	AQ939450 NRI-153R
45	91	7.8	652	9 AL646758	AL646758 AL646758

ALIGNMENTS

RESULT 1
BE857538/c
LOCUS
DEFINITION
7g01a08.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305174 3'
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGUCOSAMINE
6-O-SULFOTRANSFERASE. ; mRNA sequence.
ACCESSION BE857538
VERSION BE857538.1 GI:10371664
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS
TITLE
NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGA), Tumor Gene Index
JOURNAL
COMMENT
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seg primer: -40UP from Gibco
High quality sequence stop: 431.
Location/Qualifiers
1. .695
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3305174"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"

Db	217	ARDMLSALYRCDSLVSFQLYSPAGSGGRNLTLTGIFGAATNKVWCSSPLCPAYRKEVVG	276
QY	148	D--VCKTLCRQPFSLAREACRSYSHVVLKEVRFENLQVLYPLLSDPALNLRIVHLVRDP	205
Db	277	DDRVCCK-CPPQRLARFEEECRKYRTLVIKGVRFDVAVLAPLLRDPALDLKVIHLVRDP	335
QY	206	RAVLR-----REAGPILARD-----NGIVLGTNGKV--EADPH-LRL	242
Db	336	RAVASSRIRSRHGLIRESLQVYRSRDPRAHRMPFLEAGHKLGAKEGVGGPADIHALGA	395
QY	243	IREVCRSHVRIAEAAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLTPQLEA	302
Db	396	MEVICNSMAKTLQTA-LQPPDWLQGHYLVRYEDLVGDPVKTLRRVYDFVGLLVSPEMEQ	454
QY	303	WTHNITHGSG-IGKPIEAFTSSRNARNVSOAWRHALPFTKILRVQEVCGALQLLGYRP	361
Db	455	FALNMTSGSGSSKP---FVVSARNATQAANAWRTALTFTQIQKQVEEFCYQPMVVLGYER	511
QY	362	VYSADQQRDLTLDLV	376
Db	512	VNSPEEVKDLSKTLL	526
RESULT 13			
Q9EP78	ID	Q9EP78	PRELIMINARY; PRT; 484 AA.
AC	Q9EP78;		
DT	01-MAR-2001	(TReMBLrel. 16, Created)	
DT	01-MAR-2001	(TReMBLrel. 16, Last sequence update)	
DT	01-DEC-2001	(TReMBLrel. 19, Last annotation update)	
DE	N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE GST-5 (N-		
DE	ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE) (2600013M07RIK	PROTEIN).	
GN	CHST7 OR GST5 OR 2600013M07RIK.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa;	Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6;		
RX	PubMed=10956661;		
RA	Bhakta S., Bartes A., Bowman K.G., Kao W.M., Polsky I., Lee J.-K.,		
RA	Cook B.N., Bruehl R.E., Rosen S.D., Bertozzi C.R., Hemmerich S.;		
RT	"Sulfation of N-Acetylglucosamine by Chondroitin 6-Sulfotransferase 2		
RT	(GST-5).";		
RL	J. Biol. Chem. 275:40226-40234(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=EMBRYO;		
RX	MEDLINE=21085660; PubMed=11217851;		
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Quackenbush J.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Wagner L., Washio T.,		
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Barsh G.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,		
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,		
RA	Hayashizaki Y.;		
RT	"Functional annotation of a full-length mouse cDNA collection.";		
RL	Nature 409:685-690(2001).		
DR	EMBL; AF280089; AAG48247.1; -.		
DR	EMBL; AB040710; BAB13769.1; -.		
DR	EMBL; AK011202; BAB27465.1; -.		
DR	MGD; MGI:1891767; Chst7.		
KW	Transferase.		
SQ	SEQUENCE 484 AA; 54766 MW; 9B195537D7AB7193 CRC64;		
Query Match 30.1%; Score 617.5; DB 11; Length 484;			
Best Local Similarity 39.1%; Pred. No. 9.1e-50;			
Matches 147; Conservative 54; Mismatches 126; Indels 49; Gaps 11;			
QY	30	PGPSSPAG---GEDRVHVLVLSSWRSGSFLGQLFSQHPDVFYLMEPAWHVWTTLSQGSA	86
Db	85	PGNLSAVGEAVTQEKQHYYVHATWRTGSSFLGELFNQHPDVFYLYEPMWHLWQALYPGDA	144
QY	87	ATLHMVAVRDLMRSLFCLDMVDVFDAY-----MFQSRNLSA--FFNWATSRALCSPPA	135
Db	145	ESLQGALRDMRLSLFRCDFSVLRLYAQPGDPGERAPDSANLTAMLFRWRTNKVICSPPL	204
QY	136	CSAFFRGT----ISKQDVCKTLCRQPFSL-AREA-CRSYSHVVLKEVRFENLQVLYPLL	189
Db	205	CPAAPRARADVGLVEDKACESTC--PPVSLRALBAEACRKYPVVVIKDVRLDLLGVLYPLL	262
QY	190	SDPALNLRIVHLVRDPRAVLRSREAAGPILARDNGIVLGTNGK-----WVEADP	238
Db	263	RDPGLNLKVQQLFRDPRAVHNSRLKSRQGLLRRESIQVLRTRQRGDHFHRLVLAHGVDARP	322
QY	239	--HLRLIREVCRSHVRIAEEA-----TLKPPPEFLRGYRLVRFEDLAREPLA	283
Db	323	GGQARALPSAPRADFFLTSALEVICEAWLRDLLFTTRGAPAWLRRYLRRLRYEDLVWQPQA	382
QY	284	EIRALYAFTGLTLTPQLEAWIHNTHGSGIGKPIEAFTSSRNARNVSOAWRHALPFTKI	343
Db	383	QURRLRLRFSGLRTLAAIDAPAFNMTRGSAYGAD-RPFHLSARDAREAVHWRERLSQEQV	441
QY	344	LRVQEVCGALQLLGY	359
Db	442	RQVETACAPAMRLLAY	457
RESULT 14			
Q99NB0	ID	Q99NB0	PRELIMINARY; PRT; 484 AA.
AC	Q99NB0;		
DT	01-JUN-2001	(TReMBLrel. 17, Created)	
DT	01-JUN-2001	(TReMBLrel. 17, Last sequence update)	
DT	01-DEC-2001	(TReMBLrel. 19, Last annotation update)	
DE	CHONDROITIN 6-SULFOTRANSFERASE-2.		
GN	CHST7 OR MC6ST-2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=MAMMARY GLAND;		
RA	Kitagawa H., Uyama T., Sugahara K.;		
RT	"Cloning and Expression of Mouse Chondroitin 6-sulfotransferase-2.";		
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB046929; BAB40372.1; -.		
DR	MGD; MGI:1891767; Chst7.		
KW	Transferase.		
SQ	SEQUENCE 484 AA; 54751 MW; 402C5E1ED185FDF8 CRC64;		
Query Match 30.0%; Score 614.5; DB 11; Length 484;			
Best Local Similarity 38.8%; Pred. No. 1.8e-49;			
Matches 146; Conservative 55; Mismatches 126; Indels 49; Gaps 11;			


```

RL Genomics 55:345-347(1999).
DR EMBL; AF083066; AAD20981.1; -.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57857 MW; A82CA227B9D5651B CRC64;

Query Match
Best Local Similarity 32.2%; Score 659.5; DB 4; Length 530;
Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

QY 31 GPSSPAG-----GEDRVHVLVLSWRSRGSSFLGQLFSQHPDVFYLMPEAWHVVWTTL 81
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
145 GVAAPPNGTGTGGVGDKRQLVYVFTTWRSRGSSFFGELFNQNPVEVFFLYEPVHVWQKL 204
QY 82 SQGSAATLHMAVRDLMRISFLCDMDVFDAYMPQ---SRNLS--AFFNWAISRALCSPAC 136
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
205 YPGDAVSLQGAARDMLSALYRCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLC 264
QY 137 SAFFRGTTISKQD--VCKTLCTRQPPFSLAREACRSYSHVVLKEVRFNQLVPLSPAL 194
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
265 PAYRKEVGLVDVRVCKK-CPPQRLARFEEECRKYRLVIKGVRFDAVLAPLLEDPAL 323
QY 195 NLRIVHLVRDPRAVLR-----REAAGPILARD-----NGIVLGTNGKWV 234
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
324 DLKVIHLVRDPRAVASSRIRSRHGLIRESLQVVRSDPRAHRMPFLEAAGHKLAKKEGV 383
QY 235 --EADPH-LRLIREVCRSHVRIAEAAATLKPPPELRGRYLRVRFEDLAREPLAEIRALYAF 291
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
384 GGPADYHALGAMVEICNSMAKTLQTA-LQPPDWLQGHYLVVRYEDLVGDPVKTLRRVYDF 442
QY 292 TGLTLTPOLEAWTHNITHGSG-IGKPIEAFHTSSRNARNVSOAHRHALPFTKILRVQEV 350
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
443 VGLLVSPEMEQLPALNMTSGSGSSSKP---FVVSARNATQAANAWRTALTFOQIKQVEEFC 499
QY 351 AGALQLLGYRVPVYSADQQRDLTLDLV 376
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
500 YQPMAVLGYERVNSPEEVKDLSKTLL 525

RESULT 11
Q9UED5 PRELIMINARY; PRT; 484 AA.
AC Q9UED5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE (GLCNAC6ST).
GN GN6ST.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.";
RL J. Biochem. 124:670-678(1998).
DR EMBL; AB014679; BAA34265.1; -.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 484 AA; 53641 MW; D077EBCD7645F2F0 CRC64;

Query Match
Best Local Similarity 32.0%; Score 656.5; DB 4; Length 484;
Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;
```

```

Best Local Similarity 39.7%; Pred. No. 1.9e-53;
Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;

QY 34 SPAGGEDRVH-VLVLSWRSRGSSFLGQLFSQHPDVFYLMPEAWHVVWTTLSQGSAATLHMA 92
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
110 APEGVGDKRHHWYVFTTWRSRGSSFFGELFNQNPVEVFFLYEPVHVWQKLYPGDAVSLQGA 169
QY 93 VRDLMRISFLCDMDVFDAYMPQ---SRNLS--AFFNWAISRALCSPACSAFFRGTTISKQ 147
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
170 ARDMLSALYRCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGLV 229
QY 148 D--VCKTLCTRQPPFSLAREACRSYSHVVLKEVRFNQLVPLSPALNLRIVHLVRDP 205
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
230 DDRVCKK-CPPQRLARFEEECRKYRLVIKGVRFDAVLAPLLEDPALDKVIHLVRDP 288
QY 206 RAVLR-----REAAGPILARD-----NGIVLGTNGKWV--EADPH-LRL 242
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
289 RAVASSRIRSRHGLIRESLQVVRSDPRAHRMPFLEAAGHKLAKKEGGVPADYHALGA 348
QY 243 IREVCRSHVRIAEAAATLKPPPELRGRYLRVRFEDLAREPLAEIRALYAFGLTLPQLEA 302
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
349 MEVICNSMAKTLQTA-LQPPDWLQGHYLVVRYEDLVGDPVKTLRRVYDFVGLLVSPEME 407
QY 303 WTHNITHGSG-IGKPIEAFHTSSRNARNVSOAHRHALPFTKILRVQEVCGALQLLGYR 361
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
408 PALNMTSGSGSSSKP---FVVSARNATQAANAWRTALTFOQIKQVEEFCYQPMAVLGYER 464
QY 362 VYSADQQRDLTLDLV 376
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
465 VNSPEEVKDLSKTLL 479

RESULT 12
Q9Y4C5 PRELIMINARY; PRT; 531 AA.
AC Q9Y4C5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE LONG FORM OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE
DE (GLCNAC6ST).
GN GN6ST.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.";
RL J. Biochem. 124:670-678(1998).
DR EMBL; AB014680; BAA34266.1; -.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 531 AA; 58768 MW; D2EDB74E95B5162F CRC64;

Query Match
Best Local Similarity 32.0%; Score 656.5; DB 4; Length 531;
Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;
```

```

QY 34 SPAGGEDRVH-VLVLSWRSRGSSFLGQLFSQHPDVFYLMPEAWHVVWTTLSQGSAATLHMA 92
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
157 APEGVGDKRHHWYVFTTWRSRGSSFFGELFNQNPVEVFFLYEPVHVWQKLYPGDAVSLQGA 216
QY 93 VRDLMRISFLCDMDVFDAYMPQ---SRNLS--AFFNWAISRALCSPACSAFFRGTTISKQ 147
```


QY	301	EAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCAQALQLLGYR	360
Db	300	EAWIHNITHGSGPGARREAFKTISSRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYR	359
QY	361	PVYSADQQRDLTLDLVLPRGPDHFSWAS	388
Db	360	PVYSEDEQRNLALDLVLPRLNGFTWAS	387
RESULT	4		
Q9QUP4			
ID	Q9QUP4	PRELIMINARY;	PRT; 395 AA.
AC	Q9QUP4;		
DT	01-MAY-2000 (TReMBLrel. 13, Created)		
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)		
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)		
DE	N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE.		
GN	CHST5 OR I-GLCNAC-6-ST.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6; TISSUE=INTESTINE;		
RX	MEDLINE=99423499; PubMed=10491328;		
RA	Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;		
RT	"Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-		
RT	Sulfotransferase that is Highly Restricted to Intestinal Tissue.";		
RL	Biochem. Biophys. Res. Commun. 263:543-549(1999).		
DR	EMBL; AF176841; AAD56003.1; -.		
DR	EMBL; AF176840; AAD56002.1; -.		
DR	MGD; MGI:1931825; Chst5.		
KW	Transferase.		
SQ	SEQUENCE 395 AA; 44537 MW; 3FDF71E43ED383BE CRC64;		
		Query Match 74.9%; Score 1536.5; DB 11; Length 395;	
		Best Local Similarity 76.0%; Pred. No. 1.5e-136;	
		Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;	
QY	1	MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPAGGEDRVHVLVLSWSRSGSFLGQL	60
Db	1	MRLPRFSSTVMSLLMVQTG-ILVELVSRQVPSPAGLGERVHVLVLSWSRSGSFEVQQL	59
QY	61	FSQHPDVFYLMEPAWHVWTTLSQGSAAATLHMAVRDLMRSIFLCMDVDFDAYMPOSRNL	120
Db	60	FSQHPDVFYLMEPAWHVWTTLSQGSAPALHMAVRDLRSVFLCMDVDFDAYLPWRN	119
QY	121	FFNWATSRALCSPPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVR	180
Db	120	LFQWAVSRALCSPPVCEAFAGNISSEEVCKPLCATRPFGLAQEACSSYSHVVLKEVR	179
QY	181	NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL	240
Db	180	NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQAKALARDNGIVLGTNGTWEADPRL	239
QY	241	RLIREVCRSHVRIAEATLKPPPELGRYRLVRFEPLAREPLAEIRALYAFTGLTLPQL	300
Db	240	RVNVEVCRSHVRIAEALHKPPPELQDRYRLVRYEDLARDPLTVIRELYAFTGLTLPQL	299
QY	301	EAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCAQALQLLGYR	360
Db	300	QTIWNIHNITHGSGPGARREAFKTISSRDALSVSQAWRHLPFAKIRRVQELCGALQLLGYR	359
QY	361	PVYSADQQRDLTLDLVLPRGPDHFSWAS	388
Db	360	SVHSELEQRDLSLDLLPRGMDSFKWAS	387
RESULT	5		
Q9Y5R3			
ID	Q9Y5R3	PRELIMINARY;	PRT; 386 AA.

AC	Q9Y5R3;		
DT	01-NOV-1999 (TReMBLrel. 12, Created)		
DT	01-NOV-1999 (TReMBLrel. 12, Last sequence update)		
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)		
DE	N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE (L-SELECTIN LIGAND		
DE	SULFOTRANSFERASE GST-3).		
GN	GST3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=TONSIL;		
RX	MEDLINE=99264336; PubMed=10330415;		
RA	Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,		
RA	Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;		
RT	"Sulfotransferases of two specificities function in the reconstitution		
RT	of high endothelial cell ligands for L-selectin.";		
RL	J. Cell Biol. 145:899-910(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=TONSIL;		
RA	Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,		
RA	Rosen S.D.;		
RT	"Chromosomal Localization and Genomic Organization for the		
RT	Galactose/N-Acetylgalactosamine/N-Acetylglucosamine 6-O-		
RT	Sulfotransferase Gene Family.";		
RL	Glycobiology 0:0-0(2001).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21332592; PubMed=11439191;		
RA	Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Ellies L.G.,		
RA	Rabuka D., Hindsgaul O., Marth J.D., Lowe J.B., Fakuda M.;		
RT	"Novel sulfated lymphocyte homing receptors and their control by a		
RT	corel extension betal,3-N-acetylglucosaminyltransferase.";		
RL	Cell 105:957-969(2001).		
DR	EMBL; AF131235; AAD33015.1; -.		
DR	EMBL; AF280088; AAG48246.1; -.		
DR	EMBL; AF149783; AAK48417.1; -.		
DR	InterPro; IPR003015; HLH_MYC.		
DR	PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.		
KW	Transferase; Lectin; Selectin.		
SQ	SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;		
		Query Match 49.1%; Score 1008; DB 4; Length 386;	
		Best Local Similarity 52.3%; Pred. No. 9.9e-87;	
		Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;	
QY	14	LLLAQTTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSWSRSGSFLGQ	59
Db	1	MLLPKMKLLLLFLVSQMAILALFFHMYSHNISLSMKAQPERMHVLVLSWSRSGSFFVQ	60
QY	60	LFSQHPDVFYLMEPAWHVWTTLSQGSAAATLHMAVRDLMRSIFLCMDVDFDAYM-PQSRNL	118
Db	61	LFGQHPDVFYLMEPAWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPPRRQ	120
QY	119	SAFFNWATSRALCSPPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVR	178
Db	121	SSLEQWNSRALCSAPACDIIPQDELIIPRAHCRLLCSQQPFVEVEKACRSYSHVVLKEVR	180
QY	179	FFNLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAAAGPILARDNGIVLGTN-GKWEAD	237
Db	181	FFNLQSLYPLIKDPSLNHLIVHLVRDPRAVFRSRERTKGDLMDSRIVMGQHEQKLKED	240
QY	238	PHLRLIREVCRSHVRIAEAAATLKP-PPFLRGRYRLVRFEPLAREPLAEIRALYAFTGLTL	296
Db	241	QPYVNVQVICQSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMIEFVGLEF	298
QY	297	TPQLEAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCAQALQL	356
Db	299	LPHLQTWVHNITRGKMGD--HAFHTNARDALNVSQAWRWSLPYEKVSRLQKACGDAMNL	356

QY	121	FFNWATSRALCSPACSAFPRGTISKQDVCKTLC	TRQPFSLAREACRSYSHVVLKEVRF	180
Db	121	FFNWATSRALCSPACSAFPRGTISKQDVCKTLC	TRQPFSLAREACRSYSHVVLKEVRF	180
QY	181	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREAAGPILARDNGIVLGTNGKWEADPHL	240
Db	181	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREAAGPILARDNGIVLGTNGKWEADPHL	240
QY	241	RLIREVCRSHVRIAEAAATLKPPFLRGRYRLVRFEDLAREPLAEIRALYAFTGLTLPQL		300
Db	241	RLIREVCRSHVRIAEAAATLKPPFLRGRYRLVRFEDLAREPLAEIRALYAFTGLTLPQL		300
QY	301	EAWIHNIHSGIGCKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEV	CAGALQLLGYR	360
Db	301	EAWIHNIHSGIGCKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEV	CAGALQLLGYR	360
QY	361	PVYSADQQRDLTLDIVLP	RGPDHFSWASPD	390
Db	361	PVYSADQQRDLTLDIVLP	RGPDHFSWASPD	390
RESULT 2				
Q9GZS9				
ID	Q9GZS9	PRELIMINARY;	PRT;	411 AA.
AC	Q9GZS9;			
DT	01-MAR-2001 (TReMBLrel. 16, Created)			
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)			
DT	01-OCT-2001 (TReMBLrel. 18, Last annotation update)			
DE	INTESTINAL GLCNAC-6-SULFOTRANSFERASE (INTESTINAL N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE).			
DE	CHST5.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC	NCBI_TaxID=9606;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=20472330; PubMed=11017086;			
RX	Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,			
RA	Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,			
RA	Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,			
RA	Fukuda M.N.;			
RT	"Macular corneal dystrophy type I and type II are caused by distinct			
RT	mutations in a new sulphotransferase gene.";			
RL	Nat. Genet. 26:237-241(2000).			
DR	EMBL; AF246718; AAG28023.1; -.			
DR	EMBL; AF219991; AAG26326.1; -.			
KW	Transferase.			
SQ	SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;			
Query Match 100.0%; Score 2051; DB 4; Length 411;				
Best Local Similarity 100.0%; Pred. NO. 4.5e-185;				
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	MWLPRFSKTVTVLLLAQTTCLLLFIISRPGSPAGGEDRVHVLVLSWRSGSFLGQL	60	
Db	22	MWLPRFSKTVTVLLLAQTTCLLLFIISRPGSPAGGEDRVHVLVLSWRSGSFLGQL	81	
QY	61	FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMRSIFLCMDVDFDAYMPOSRLSA	120	
Db	82	FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMRSIFLCMDVDFDAYMPOSRLSA	141	
QY	121	FFNWATSRALCSPACSAFPRGTISKQDVCKTLC	TRQPFSLAREACRSYSHVVLKEVRF	180
Db	142	FFNWATSRALCSPACSAFPRGTISKQDVCKTLC	TRQPFSLAREACRSYSHVVLKEVRF	201
QY	181	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREAAGPILARDNGIVLGTNGKWEADPHL	240
Db	202	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREAAGPILARDNGIVLGTNGKWEADPHL	261
QY	241	RLIREVCRSHVRIAEAAATLKPPFLRGRYRLVRFEDLAREPLAEIRALYAFTGLTLPQL	300	

Db	262	RLIREVCRSHVRIAEAAATLKPPPEFLGRYRLVRFEDLAREPLAEIRALYAFTGLTLPQL	321
QY	301	EAWIHNIHSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEV	CAGALQLLGYR 360
Db	322	EAWIHNIHSGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEV	CAGALQLLGYR 381
QY	361	PVYSADQQRDLTLDVLP	RGPDHFSWASPD 390
Db	382	PVYSADQQRDLTLDVLP	RGPDHFSWASPD 411
RESULT 3			
ID	Q9GZX3	PRELIMINARY;	PRT; 395 AA.
AC	Q9GZX3;		
DT	01-MAR-2001 (TReMBLrel. 16, Created)		
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)		
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)		
DE	N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE GST-4BETA (CORNEAL N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE).		
DE	GST4BETA OR CHST6.		
GN	Homo sapiens (Human).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OC	NCBI_TaxID=9606;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R., Rosen S.D.;		
RA	"Chromosomal Localization and Genomic Organization for the Galactose/N-Acetylgalactosamine/N-Acetylglucosamine 6-O-Sulfotransferase Gene Family.";		
RT	Glycobiology 0:0-0(2001).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=20472330; PubMed=11017086;		
RX	Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T., Nakamura T.,		
RA	Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S., Ozaki K.,		
RA	Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;		
RT	"Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene.";		
RL	Nat. Genet. 26:237-241(2000).		
DR	EMBL; AF280086; AAG48244.1; -.		
DR	EMBL; AF219990; AAG26325.1; -.		
DR	EMBL; AF219991; AAG26327.1; -.		
KW	Transferase.		
SQ	SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;		
Query Match 84.3%; Score 1729.5; DB 4; Length 395;			
Best Local Similarity 85.8%; Pred. No. 9.1e-155;			
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;			
QY	1	MWLPRFSKTVTVLLLAQTTCLLLFIISRP	GSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
Db	1	MWLPRVSSFAVTALLAQ-TFLLLFVSR	PGPSSPAGGEARVHVLVLSWRSGSFGQL 59
QY	61	FSQHPDVFYLMPEPAWHVWTTLSQGS	AATLHMAVRDLMRSIFLCMDVDFDAYMPOSRLSA 120
Db	60	FNQHPDVFYLMPEPAWHVWTTLSQGS	AATLHMAVRDLVRSVFLCDMDVDFDAYLPWRRNLS 119
QY	121	FFNWATSRALCSPACSAFPRGTISKQDV	CKTLC
Db	120	LFQWAVSRALCSPACSAFPRGAISS	EAACKPLCARQSFILAREACRSYSHVVLKEVRF 179
QY	181	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREAAGPILARDNGIVLGTNGKWVEADPHL 240
Db	180	NLQVLYPLSDPALNLRIVHLVRDPRAVLR	SREQTAKALARDNGIVLGTNGTWEADPGL 239
QY	241	RLIREVCRSHVRIAEAAATLKPPPF	FLGRYRLVRFEDLAREPLAEIRALYAFTGLTLPQL 300
Db	240	RVREVC	SHVRIAEAAATLKPPPF

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 19:58:11 ; Search time 32.28 Seconds
(without alignments)
2090.087 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPFRSSKTVTVLLAQT.....LTLDLVLPRGPDHFSWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_19:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2051	100.0	390	4 Q9UBY3	Q9uby3 homo sapien
2	2051	100.0	411	4 Q9GZS9	Q9gzs9 homo sapien
3	1729.5	84.3	395	4 Q9GZX3	Q9gzx3 homo sapien
4	1536.5	74.9	395	11 Q9QUP4	Q9qup4 mus musculu
5	1008	49.1	386	4 Q9Y5R3	Q9y5r3 homo sapien
6	965.5	47.1	388	11 Q9R1I1	Q9rl1i mus musculu
7	958.5	46.7	388	11 Q9WUE5	Q9wue5 mus musculu
8	665	32.4	530	11 Q88276	Q88276 mus musculu
9	659.5	32.2	483	4 Q9GZN5	Q9gzn5 homo sapien
10	659.5	32.2	530	4 Q9Y6F2	Q9y6f2 homo sapien
11	656.5	32.0	484	4 Q9UED5	Q9ued5 homo sapien
12	656.5	32.0	531	4 Q9Y4C5	Q9yac5 homo sapien
13	617.5	30.1	484	11 Q9EP78	Q9ep78 mus musculu
14	614.5	30.0	484	11 Q99NB0	Q99nb0 mus musculu
15	610	29.7	486	4 Q75667	Q75667 homo sapien
16	609	29.7	486	4 Q9NS84	Q9ns84 homo sapien

17	577.5	28.2	479	4 Q75099	Q75099 homo sapien
18	576	28.1	472	11 Q88199	Q88199 mus musculu
19	560	27.3	474	11 Q9QZL2	Q9qzl2 rattus norv
20	520	25.4	411	11 Q9EQC0	Q9eqc0 mus musculu
21	513.5	25.0	411	4 Q43916	Q43916 homo sapien
22	472.5	23.0	441	13 Q93403	Q93403 torpedo cal
23	274	13.4	486	5 Q95TN9	Q95tn9 drosophila
24	197	9.6	119	6 Q95JA8	Q95ja8 oryctolagus
25	184.5	9.0	183	5 Q9VMC2	Q9vmc2 drosophila
26	174.5	8.5	363	5 Q9VMC4	Q9vmc4 drosophila
27	151	7.4	114	11 Q9DOK5	Q9d0k5 mus musculu
28	134	6.5	303	2 Q93JE6	Q93je6 streptomyce
29	107.5	5.2	307	16 Q92VG4	Q92vg4 rhizobium m
30	107.5	5.2	315	5 Q9VMC3	Q9vmc3 drosophila
31	99.5	4.9	972	5 Q9VLX0	Q9vlix0 drosophila
32	95	4.6	474	3 Q14192	Q14192 schizosacch
33	95	4.6	1189	16 Q92TC4	Q92tc4 rhizobium m
34	94	4.6	582	16 Q53639	Q53639 mycobacteri
35	94	4.6	933	5 Q9GNZ1	Q9gnz1 leishmania
36	93.5	4.6	615	16 Q9RRS5	Q9rrs5 deinococcus
37	93	4.5	523	16 Q9I6Z6	Q9i6z6 pseudomonas
38	92	4.5	925	17 Q9HLE8	Q9hle8 thermoplas
39	91	4.4	920	16 Q9A0U7	Q9a0u7 streptococc
40	89.5	4.4	1844	12 Q56254	Q56254 turnip yell
41	88.5	4.3	332	2 Q939Q8	Q939q8 streptomyce
42	88	4.3	417	2 Q83044	Q83044 azospirillu
43	87.5	4.3	350	10 Q94DZ3	Q94dz3 oryza sativ
44	87.5	4.3	661	5 Q9NKE9	Q9nke9 drosophila
45	87.5	4.3	664	5 Q9VJW2	Q9vjw2 drosophila

ALIGNMENTS

RESULT 1

Q9UBY3 ID Q9UBY3 PRELIMINARY; PRT; 390 AA.

AC Q9UBY3;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)

DE N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE.

GN I-GLCNAC-6-St.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=INTESTINE;

RX MEDLINE=9423499; PubMed=10491328;

RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;

RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-Sulfotransferase that is Highly Restricted to Intestinal Tissue."

RL Biochem. Biophys. Res. Commun. 263:543-549(1999).

DR EMBL; AF176839; AAD56001.1; -.

DR EMBL; AF176838; AAD56000.1; -.

KW Transferase.

SQ SEQUENCE 390 AA; 43832 MW; 132D1C3B6BE38179 CRC64;

Query Match 100.0%; Score 2051; DB 4; Length 390;
Best Local Similarity 100.0%; Pred. No. 4.2e-185;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MWLPFRSSKTVTVLLAQTTCLLFIISRPGPSPPAGGEDRVHVLVLSWSRSGSFLGQL	60
Db	1	MWLPFRSSKTVTVLLAQTTCLLFIISRPGPSPPAGGEDRVHVLVLSWSRSGSFLGQL	60
QY	61	FSQHPDVFYLMPEPAWHVWTTLSQSSAATLHMAVRDLMRSIFLCMDVDFDAYMPQSRNLSA	120
Db	61	FSQHPDVFYLMPEPAWHVWTTLSQSSAATLHMAVRDLMRSIFLCMDVDFDAYMPQSRNLSA	120